"Nonparametric incidence and latency estimation in mixture cure models"

López-Cheda , Ana ; Cao, Ricardo ; Jacome, María Amalia ; Van Keilegom, Ingrid

Abstract

A completely nonparametric method for the estimation of mixture cure models is proposed in this paper. The nonparametric estimator of the incidence introduced by Xu and Peng (2014) is extensively studied and a nonparametric estimator of the latency is presented. These estimators, which are based on the Beran estimator of the conditional survival function, are proved to be the local maximum likelihood estimators. An iid representation is obtained for the nonparametric incidence estimator. As a consequence, an asymptotically optimal bandwidth is found. Moreover, a bootstrap bandwidth selection method for the nonparametric incidence estimator is proposed. The introduced nonparametric estimators are compared with existing semiparametric approaches in a simulation study, in which the performance of the bootstrap bandwidth selector is also assessed. Finally, the presented method is applied to a database of colorectal cancer from the University Hospital of A Coruña (CHUAC).

Document type : *Document de travail (Working Paper)*

I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2015/14

Nonparametric incidence and latency estimation in mixture cure models

LÓPEZ-CHEDA, A., CAO, R., JÁCOME, M.A. AND I. VAN KEILEGOM

Nonparametric incidence and latency estimation in mixture cure models

A. López-Cheda[1], R. Cao[1], M. A. Jácome[1] and I. Van Keilegom[2]

[1] *University of A Coruña; ana.lopez.cheda@udc.es, rcao@udc.es, majacome@udc.es*
[2] *Université catholique de Louvain; ingrid.vankeilegom@uclouvain.be*

Abstract

A completely nonparametric method for the estimation of mixture cure models is proposed in this paper. The nonparametric estimator of the incidence introduced by Xu and Peng (2014) is extensively studied and a nonparametric estimator of the latency is presented. These estimators, which are based on the Beran estimator of the conditional survival function, are proved to be the local maximum likelihood estimators. An iid representation is obtained for the nonparametric incidence estimator. As a consequence, an asymptotically optimal bandwidth is found. Moreover, a bootstrap bandwidth selection method for the nonparametric incidence estimator is proposed. The introduced nonparametric estimators are compared with existing semiparametric approaches in a simulation study, in which the performance of the bootstrap bandwidth selector is also assessed. Finally, the presented method is applied to a database of colorectal cancer from the University Hospital of A Coruña (CHUAC).

Key words: Survival analysis, censored data, maximum likelihood, kernel estimation, bootstrap bandwidth selector

1 Introduction

Thanks to the effectiveness of current cancer treatments, the proportion of patients who get cured (or who at least survive for a long time) is increasing over time. Therefore, data coming from cancer studies typically have heavy censoring at the end of the follow-up period, and a standard survival model is inappropriate. To accommodate for the cured or insusceptible proportion of subjects, a cure fraction can be explicitly incorporated into survival models and, as a consequence, cure models arise. These models allow to estimate the cured proportion (incidence) and also the probability of survival of the uncured patients up to a given time point (latency). Note that cure models should not be used indiscriminately (Farewell, 1986), there must be good empirical and biological evidence of a insusceptible population.

There are two main classes of cure models: mixture and non-mixture models. The first paper in non-mixture models was due to Haybittle (1959, 1965). One category, belonging to this group, is the proportional hazards (PH) cure model, also known as the promotion time cure model, first proposed by Yakovlev and Tsodikov (1996). The parameters

in this model can be estimated parametrically (Yakovlev et al, 1994; Chen et al, 1999; Chen et al, 2002) or semiparametrically (Tsodikov, 1998, 2003; Zeng et al., 2006). Moreover, Tsodikov (2001) proposed a nonparametric estimator of the incidence, but it cannot handle continuous covariates.

In this paper we consider a model which belongs to the other category of cure models, called two-component mixture cure models. The mixture cure model was proposed by Boag (1949) and it explicitly expresses the survival function as a mixture of two types of patients: those who are cured and those who are not. An advantage of this model is that it allows the covariates to have different influence on cured and uncured patients. Maller and Zhou (1996) provided a detailed review of this model. In mixture cure models, the incidence is usually assumed to have a logistic form and the latency is usually estimated parametrically (Jones et al., 1981; Farewell, 1982, 1986; Cantor and Shuster, 1992; Ghitany et al., 1994; Denham et al., 1996) or semiparametrically (Kuk and Chen, 1992; Yamaguchi, 1992; Peng et al., 1998; Peng and Dear, 2000; Sy and Taylor, 2000; Li and Taylor, 2002; Zhang and Peng, 2007).

Due to the fact that the effects of the covariate on the cure rate cannot always be well approximated using parametric or semiparametric methods, a nonparametric approach is needed. In the literature, some nonparametric methods for the estimation of the cure rate have been studied: Maller and Zhou (1992) proposed a consistent nonparametric estimator of the incidence, but it cannot handle covariates. In order to overcome this drawback, Laska and Meisner (1992) proposed another nonparametric estimator of the cure rate, but it only works for discrete covariates. Furthermore, Wang et al. (2012) proposed a cure model with a nonparametric form in the cure probability. To ensure model identifiability, they assumed a nonparametric proportional hazards model for the hazard function. The estimation was carried out by an expectation-maximization algorithm for a penalized likelihood. They defined the smoothing spline function estimates as the minimizers of the penalized likelihood. Although the above papers have a nonparametric flavor, they fail to consider a completely nonparametric mixture cure model which works for discrete and continuous covariates in both the incidence and the latency. More recently, Xu and Peng (2014) extended the existing work by proposing a nonparametric incidence estimator which allows for a continuous covariate. The present paper will study that nonparametric incidence estimator deeply and fill this important gap for the latency function.

In this paper, we propose a two-component mixture model with nonparametric forms for both the cure probability and the survival function of the uncured individuals. Identifiability of the model is guaranteed by considering the time beyond which we have no interest as infinity and by identifying censored subjects beyond the largest observable failure time as cured, so that our estimator does not overestimate the true cure rate.

The rest of the article is organized as follows. In Section 2 we give a detailed description of our nonparametric mixture cure model, we study the estimator of the incidence proposed in Xu and Peng (2014) and we introduce a nonparametric estimator of the latency. Moreover, we address the identifiability problem. We also present a local maximum likelihood result as well as an iid representation and the asymptotic mean squared error for the nonparametric incidence estimator. A bootstrap bandwidth selection method is introduced in Section 3. Section 4 includes a comparison between these nonparametric estimators and the semiparametric ones proposed in Peng and Dear (2000) in a simulation study and assesses the practical performance of the bootstrap bandwidth selector. In Section 5 we apply the proposed nonparametric method to real data related to colorectal cancer

patients in CHUAC. An appendix contains the proofs.

2 Nonparametric mixture cure model

2.1 Notation

Let ξ be a binary variable where $\xi = 0$ indicates if the individual belongs to the susceptible group (the individual will eventually experience the event of interest if followed for long enough) and $\xi = 1$ indicates if the subject is cured (the individual will never experience the event). The proportion of cured patients and the survival function in the group of uncured patients can depend on certain characteristics of the subject, represented by a set of covariates X. Let $p(x) = P(\xi = 0|X = x)$ be the conditional probability of not being cured, and let Y be the time to occurrence of the event. When $\xi = 1$ it is assumed that $Y = \infty$.

The conditional distribution function of Y is $F(t|x) = P(Y \leq t|X = x)$. Note that the corresponding survival function, $S(t|x)$, is improper when cured patients exist, since $\lim_{t\to\infty} S(t|x) = 1 - p(x) > 0$. The conditional survival function of Y given that the subject is not cured is denoted by $S_0(t|x) = P(Y > t|X = x, \xi = 0)$. Then, the mixture cure model can be written as:

$$S(t|x) = 1 - p(x) + p(x)S_0(t|x), \tag{1}$$

where $1-p(x)$ is the incidence and $S_0(t|x)$ is the latency. We assume that each individual is subject to random right censoring and that the censoring time, C, with distribution function G, is independent of Y given the covariates X. Let $T = \min(Y, C)$ be the observed time with distribution function H and $\delta = I(Y \leq C)$ the uncensoring indicator. Observe that $\delta = 0$ for all the cured patients, and it also happens for uncured patients with censored lifetime. Let X be a univariate continuous covariate with density function $m(x)$. Therefore, the observations will be $(X_i, T_i, \delta_i), i = 1, \ldots, n$ independent and identically distributed copies of the random vector (X, T, δ).

In order to introduce the nonparametric approach in mixture cure models, we consider the generalized Kaplan-Meier estimator by Beran (1981) to estimate the conditional survival function with covariates:

$$\hat{S}_h(t|x) = \prod_{T_i \leq t} \left(1 - \frac{\delta_{(i)} B_{h(i)}(x)}{\sum_{r=i}^{n} B_{h(r)}(x)}\right), \tag{2}$$

where

$$B_{h(i)}(x) = \frac{K_h(x - X_{(i)})}{\sum_{j=1}^{n} K_h(x - X_{(j)})} \tag{3}$$

are the Nadaraya-Watson (NW) weights with $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ the rescaled kernel with bandwidth $h \to 0$. Here $T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(n)}$ are the ordered T_i's, and $\delta_{(i)}$ and $X_{(i)}$ are the corresponding uncensoring indicator and covariate concomitants. We will also denote $\hat{F}_h(t|x) = 1 - \hat{S}_h(t|x)$ the Beran estimator of $F(t|x)$.

Departing from the Beran estimator, Xu and Peng (2014) introduced the following kernel

type estimator of the incidence:

$$1-\hat{p}_h(x)=\prod_{i=1}^{n}\left(1-\frac{\delta_{(i)}B_{h(i)}(x)}{\sum_{r=i}^{n}B_{h(r)}(x)}\right)=\hat{S}_h(T^1_{\max}|x), \tag{4}$$

where $T^1_{\max}=\max_{i:\delta_i=1}(T_i)$ is the largest uncensored failure time. Using (1) it is straightforward to prove that

$$S_0(t|x)=\frac{S(t|x)-(1-p(x))}{p(x)}.$$

As a consequence, we propose the following nonparametric estimator of the latency:

$$\hat{S}_{0,h}(t|x)=\frac{\hat{S}_h(t|x)-(1-\hat{p}_h(x))}{\hat{p}_h(x)}, \tag{5}$$

where $\hat{S}_h(t|x)$ is the Beran estimator of $S(t|x)$ in (2) and $1-\hat{p}_h(x)$ is the estimator by Xu and Peng (2014) in (4).

The identifiability of a cure model allows to obtain unique estimates of the model parameters. In a cure model, all observed uncensored subjects ($\delta_i=1$) are necessarily uncured ($\xi_i=0$); but it is impossible to say if an observed censored individual ($\delta_i=0$) belongs to the susceptible group ($\xi_i=0$) or to the non-susceptible group ($\xi_i=1$), because some censored subjects may experience failures beyond the study period. This leads to difficulties in making a distinction between models with high incidence of being susceptible and long tails of the latency distribution and low incidence of being susceptible and short tails of the latency distribution. To address this problem, we present Lemma 1.

Lemma 1 *Let T^+ be an arbitrary large time (possibly $T^+=\infty$) and let D be the support of X. Model (1), with $p(x)$ and $S_0(t|x)$ unspecified, is identifiable if T^+ is large enough such that $S_0(T^+|x)=0$ for all $x\in D$ and $t<T^+$.*

2.2 Asymptotic properties

The Beran estimator of the conditional survival function has been deeply studied in the literature. Dabrowska (1989), in Theorem 2.1, shows its asymptotic unbiasedness, considering Nadaraya-Watson weights. Furthermore, using Gasser-Müller weights, González-Manteiga and Cadarso-Suárez (1994) give an almost sure representation for the estimator as a sum of independent variables plus a remainder term, and Van Keilegom and Veraverbeke (1997a) prove an asymptotic representation for the bootstrapped estimator and obtain a strong consistency of the bootstrap approximation for the conditional distribution function.

Let $\widehat{\Lambda}_h(t|x)$ be the estimator of the conditional cumulative hazard function:

$$\widehat{\Lambda}_h(t|x)=\sum_{i=1}^{n}\frac{\delta_{(i)}B_{h(i)}(x)}{\sum_{r=i}^{n}B_{h(r)}(x)}I(T_{(i)}\leq t)=\int_0^t\frac{d\hat{H}^1_h(v|x)}{1-\hat{H}_h(v^-|x)},$$

where

$$\hat{H}_h(t|x)=\sum_{i=1}^{n}B_{hi}(x)I(T_i\leq t)$$

and

$$\hat{H}_h^1(t|x) = \sum_{i=1}^{n} B_{hi}(x) I(T_i \leq t, \delta_i = 1)$$

are the empirical estimators of

$$H(t|x) = P\left(T \leq t | X = x\right) \quad \text{and} \quad H^1(t|x) = P\left(T \leq t, \delta = 1 | X = x\right)$$

respectively. If $\bar{G}(t|x) = 1 - G(t|x)$ denotes the proper conditional survival function of the censoring time C, and assuming that $S_0(t|x)$ is a proper survival function, we define:

$$\begin{aligned}
\tau_H(x) &= \sup\left\{t : H(t|x) < 1\right\}, \\
\tau_{S_0}(x) &= \sup\left\{t : S_0(t|x) > 0\right\}, \\
\tau_G(x) &= \sup\left\{t : G(t|x) < 1\right\}.
\end{aligned}$$

Since $S(t|x)$ is an improper survival function, that is, $S(t|x) > 0$ for any $t \in [0, \infty)$, and $1 - H(t|x) = S(t|x) \times \bar{G}(t|x)$, we have

$$\tau_H(x) = \tau_G(x).$$

Let $\tau_0 = \sup_{x \in D} \tau_{S_0}(x)$. As in Xu and Peng (2014), we assume

$$\tau_0 < \tau_G\left(x\right), \forall x \in D. \tag{6}$$

This condition states that the support of the censoring variable is not contained in the support of Y, which guarantees that censored subjects beyond the largest observable failure time are cured. Hence, our estimator does not overestimate the true cure rate. A similar assumption was used by Maller and Zhou (1992, 1994) in homogeneous cases. As pointed out in Laska and Meisner (1992), if the censoring variable takes values always below a time $\tau_G < \tau_0$, for example in a clinical trial with a fixed maximum follow-up period, the largest uncensored observation T_{max}^1 may occur at a time not larger than τ_G and therefore always before τ_0. In such a case, for a large sample size, the estimator in (4) is an estimator of $1 - p(x) + p(x)S_0(\tau_G)$ which is strictly larger than $1 - p(x)$. This comment shows the need of considering the length of follow-up in the design of a clinical trial carefully, so that $S_0(\tau_G)$ is sufficiently small to take the estimator (4) of $1-p(x)+p(x)S_0(\tau_G)$ as a good estimator of $1-p(x)$ for practical purposes. The simulations in Xu and Peng (2014) show that if the censoring distribution $G(t|x)$ has a heavier tail than $S_0(t|x)$, the estimates from the proposed method will tend to have smaller biases regardless of the value of $\tau_{S_0}(x)$.

We further assume that:

(A1) X, Y and C are absolutely continuous random variables.

(A2) (a) Let $I = [x_1, x_2]$ be an interval contained in the support of m, and $I_\delta = [x_1 - \delta, x_2 + \delta]$ for some $\delta > 0$ such that

$$0 < \gamma = \inf[m\left(x\right) : x \in I_\delta] < \sup[m\left(x\right) : x \in I_\delta] = \Gamma < \infty$$

and $0 < \delta\Gamma < 1$. And for all $x \in I_\delta$ the random variables Y and C are conditionally independent given $X = x$.

(b) There exist $a, b \in \mathbb{R}$, with $a < b$ satisfying $1-H(t|x) \geq \theta > 0$ for $(t, x) \in [a, b] \times I_\delta$.

(A3) The first derivative of the function $m(x)$ exists and is continuous in $x \in I_\delta$ and the first derivatives with respect to x of the functions $H(t|x)$ and $H^1(t|x)$ exist and are continuous and bounded in $(t,x) \in [0,\infty) \times I_\delta$.

(A4) The second derivative of the function $m(x)$ exists and is continuous in $x \in I_\delta$ and the second derivatives with respect to x of the functions $H(t|x)$ and $H^1(t|x)$ exist and are continuous and bounded in $(t,x) \in [0,\infty) \times I_\delta$.

(A5) The first derivatives with respect to t of the functions $G(t|x)$, $H(t|x)$, $H^1(t|x)$ and $S_0(t|x)$ exist and are continuous in $(t,x) \in [a,b] \times D$.

(A6) The second derivatives with respect to t of the functions $H(t|x)$ and $H^1(t|x)$ exist and are continuous in $(t,x) \in [a,b] \times D$.

(A7) Let us define $H_{c,1}(t) = P(T < t|\delta = 1)$. The first and second derivatives of the distribution and subdistribution functions $H(t)$ and $H_{c,1}(t)$ are bounded away from zero in $[a,b]$. Moreover, $H'_{c,1}(\tau_0) > 0$.

(A8) The kernel function K is a symmetric density vanishing outside $(-1,1)$ and the total variation of K is less than some $\lambda < \infty$.

(A9) The kernel K is a twice differentiable function with K'' continuous.

(A10) The functions $H(t|x)$, $S_0(t|x)$ and $G(t|x)$ have bounded second-order derivatives with respect to x for any given value of t.

(A11) $\int_0^\infty \frac{dH^1(t|x)}{(1-H(t|x))^2} < \infty \ \forall x \in I.$

The consistency and asymptotic normality of the incidence estimator are proved in Xu and Peng (2014). In the next theorem we show that both the proposed nonparametric incidence and latency estimators are the local maximum likelihood estimators of $1-p(x)$ and $S_0(t|x)$.

Theorem 2 *The kernel type estimators* $1-\hat{p}_h(x)$ *and* $\hat{S}_{0,h}(t|x)$, *given in (4) and (5) respectively, are the local maximum likelihood estimators of* $1-p(x)$ *and* $S_0(t|x)$ *for the mixture cure model (1), for any* $x \in D$ *and* $t \geq 0$.

We also obtain an iid representation of the incidence estimator.

Theorem 3 *Under assumptions* $(A1)-(A11)$ *and assuming (6), for a sequence of bandwidths satisfying* $nh^5(\ln n)^{-1} = O(1)$ *and* $\ln n/(nh) \to 0$, *then*

$$(1-\hat{p}_h(x)) - (1-p(x)) = (1-p(x)) \sum_{i=1}^{n} \tilde{B}_{hi}(x) \xi(T_i, \delta_i, x) + R_n(x)$$

where

$$\tilde{B}_{hi}(x) = \frac{\frac{1}{nh} K\left(\frac{x-X_i}{h}\right)}{m(x)}, \tag{7}$$

$$\xi(T_i, \delta_i, x) = \frac{I(\delta_i = 1)}{1-H(T_i|x)} - \int_0^{T_i} \frac{dH^1(t|x)}{(1-H(t|x))^2} \tag{8}$$

and

$$\sup_{x \in I} |R_n(x)| = O\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \ a.s.$$

Finally, from the representation in Theorem 3 with straightforward calculations, the asymptotic expression of the mean squared error of the incidence estimator, $MSE_x(h_x) =$

$E[(\hat{p}_{h_x}(x) - p(x))^2]$, is given by:

$$AMSE_x(h) = \frac{1}{nh}(1-p(x))^2 c_K \sigma^2(x) + \left[h^2 \frac{1}{2} d_K (1-p(x))\mu(x)\right]^2, \tag{9}$$

where the first term corresponds to the asymptotic variance and the second one to the asymptotic squared bias, with $d_K = \int v^2 K(v)dv$, $c_K = \int K^2(v)dv$ and, following a notation similar to that in Dabrowska (1992):

$$\mu(x) = \frac{2\Phi'(x,x)m'(x) + \Phi''(x,x)m(x)}{m(x)},$$

$$\sigma^2(x) = \frac{1}{m(x)} \int_0^\infty \frac{dH^1(t|x)}{(1-H(t|x))^2},$$

where

$$\Phi(u,x) = \int_0^\infty \frac{dH^1(t|u)}{1-H(t|x)} - \int_0^\infty \frac{1-H(t|u)}{(1-H(t|x))^2} dH^1(t|x),$$

with $\Phi'(u,x) = \partial/(\partial u)\Phi(u,x)$ and $\Phi''(u,x) = \partial^2/(\partial u^2)\Phi(u,x)$.

3 Bandwidth selection

The choice of the bandwidth is a crucial issue for kernel estimation, since it controls the trade-off between bias and variance. Note that the nonparametric estimator of $1-p(x)$ is a generalization of the Kaplan-Meier estimator, in which the weights n^{-1} are now replaced with more general weights $B_{hi}(x)$ that depend on the relative position of each observed value of the covariate X_i with respect to x controlled by the bandwidth h. When the bandwidth is too large, then $B_{hi}(x) \simeq n^{-1}$ and the nonparametric estimator reduces to the homogeneous cure rate estimator proposed by Maller and Zhou (1992), in which the effect of the covariate is left out, increasing the bias. On the other hand, if the bandwidth is too small, then the error of estimation shoots up at the expense of the variance. Several methods for the selection of the smoothing parameter have been proposed in the literature. They mainly look for a small error when approximating the underlying curve by the smooth estimate. The asymptotically optimal local bandwidth to estimate the cure rate, $1-p(x)$, in the sense of minimizing the asymptotic expression of the MSE_x in (9), is given by:

$$h_{x,AMSE} = \left(\frac{c_K \sigma^2(x)}{d_K^2 \mu^2(x)}\right)^{1/5} n^{-1/5}.$$

This optimal bandwidth depends on unknown underlying distributions through $\mu(x)$ and $\sigma^2(x)$. Considering Dabrowska (1989), a plug-in bandwidth selector can be obtained by replacing the unknown functions in $\mu(x)$ and $\sigma^2(x)$ by consistent nonparametric estimates computed with unknown pilot bandwidths, giving rise to a never-ending process, which seems even harder than the original problem of incidence estimation. On the other hand, unfortunately, the finite-sample behavior of the cross validation (CV) bandwidth selector in this context turned out to be disappointing. The CV bandwidth was highly variable and tended to undersmooth its kernel estimate.

Another way to select the bandwidth is to use the bootstrap method. It consists of minimizing a bootstrap estimate of $MSE_x(h_x)$, rather than minimizing the estimation of its asymptotic expression in (9).

The bootstrap method was introduced by Efron (1979) in a complete and homogeneous setup, in which the bootstrap is carried out by resampling with replacement from the sample. Bootstrap for right censored data with no covariates was first studied by Efron (1981), who proposed two equivalent bootstrap versions (simple and obvious bootstrap), Reid (1981) and Akritas (1986). Bootstrap for right censored data was later studied by Lo and Singh (1986), Horváth and Yandell (1987) and Lai and Wang (1993), among others.

We depart from the resampling proposed by Li and Datta (2001). They give two methods for bootstrapping the Beran estimate of the conditional survival function, the simple weighted bootstrap and the obvious bootstrap, and show the equivalence of both algorithms, which is parallel to the equivalence between Efron's (1981) resampling methods.

In this paper, we consider the simple weighted bootstrap, without resampling the covariate X but just fixing the covariate sample in the resamples, which is equivalent to the original simple weighted bootstrap proposed by Li and Datta (2001). For fixed x and $i = 1, \ldots, n$, we set $X_i^* = X_i$ and generate a pair (T_i^*, δ_i^*) from the weighted empirical distribution $\hat{F}_{g_x}(\cdot, \cdot | X_i^*)$, where

$$\hat{F}_{g_x}(u, v|x) = \sum_{i=1}^{n} B_{g_x i}(x) I(T_i \leq u, \delta_i \leq v)$$

and $B_{g_x i}(x)$ is the NW weight in (3) with pilot bandwidth g_x. The resulting bootstrap resample is $\{(X_1, T_1^*, \delta_1^*), \ldots, (X_n, T_n^*, \delta_n^*)\}$.

The bootstrap bandwidth is the minimizer of the bootstrap version of $MSE_x(h_x)$,

$$MSE^*_{x,g_x}(h_x) = E^*[(\hat{p}^*_{h_x,g_x}(x) - \hat{p}_{g_x}(x))^2], \tag{10}$$

that can be approximated, using Monte Carlo, by:

$$MSE^*_{x,g_x}(h_x) \simeq \frac{1}{B} \sum_{b=1}^{B} (\hat{p}^{*b}_{h_x,g_x}(x) - \hat{p}_{g_x}(x))^2, \tag{11}$$

where $\hat{p}^{*b}_{h_x,g_x}(x)$ is the kernel estimator of p using bandwidth h_x and based on the b-th bootstrap resample generated from $\hat{F}_{g_x}$, and $\hat{p}_{g_x}(x)$ is the kernel estimator of p computed with the original sample and pilot bandwidth g_x.

The procedure for obtaining the bootstrap bandwidth selector for a fixed value x of the covariate is as follows:

1. Generate B bootstrap resamples of the form $\left\{ \left(X_1^{(b)}, T_1^{*(b)}, \delta_1^{*(b)}\right), \ldots, \left(X_n^{(b)}, T_n^{*(b)}, \delta_n^{*(b)}\right) \right\}$, $b = 1, \ldots, B$.
2. For the b-th bootstrap resample ($b = 1, ..., B$), compute the nonparametric estimator $\hat{p}^{*b}_{h_l,g_x}$ with bandwidth h_l, $l = 1, 2, \ldots, L$.
3. With the original sample and pilot bandwidth g_x, compute $\hat{p}_{g_x}(x)$.
4. For each bandwidth h_l in the grid, compute the Monte Carlo approximation of $MSE^*_{x,g_x}(h_l)$ given by (11).

5. The bootstrap bandwidth, h_x^*, is the minimizer of the Monte Carlo approximation of $MSE^*_{x,g_x}(h_l)$ over the grid of bandwidths $\{h_1, \ldots, h_L\}$.

The $MSE^*_{x,g_x}(h_x)$ depends on the unknown pilot bandwidth g_x. The optimal pilot bandwidth g_x should be chosen so that it minimizes (10) for a given sample. The idea is to minimize the dominant term of an a.s. asymptotic representation of $MSE^*_{x,g_x}(h_x)$, that will not have any unknown functions to be estimated. Our arguments are based on the results in Van Keilegom and Veraverbeke (1997a,b) for the bootstrapped Beran estimator which were carried out with non-random covariates. However, the extension to random design is feasible by replacing the Gasser-Müller weights with the NW weights in (3).

Similarly as in Theorem 3 in Van Keilegom and Veraverbeke (1997b) for the bootstrapped Beran estimator $\hat{F}^*_{h_x,g_x}(\cdot|x)$, and keeping in mind that $\hat{p}^*_{h_x,g_x}(x) = \hat{F}^*_{h_x,g_x}(T^1_{\max}|x)$, with $h_x = Cn^{-1/5}$ for some $C > 0$, we have

$$\begin{aligned}(nh_x)^{1/2}(\hat{p}^*_{h_x,g_x}(x) - \hat{p}_{g_x}(x)) &= W^*_{h_x}(x) + (nh_x)^{1/2}(\hat{b}_{h_x,g_x}(x) - b_{h_x}(x)) \\ &\quad + \Big((nh_x)^{1/2} b_{h_x}(x) - b(x)\Big) + b(x) + r^*_n(x), \qquad (12)\end{aligned}$$

where

$$W^*_{h_x}(x) = (nh_x)^{1/2} \sum_{i=1}^n B_{h_x(i)}(x) \left(\xi(T^*_{(i)}, \delta^*_{(i)}, x) - E^* \left(\xi(T^*_{(i)}, \delta^*_{(i)}, x)\right)\right)$$

can be proved to converge to a zero mean Gaussian process, in a similar way as in Theorem 3.1 in Van Keilegom and Veraverbeke (1997a), using straightforward calculations and Lemma 3.2 in that paper, where every H (and H^1) is replaced with H_{g_x} (and $H^1_{g_x}$),

$$\begin{aligned}\hat{b}_{h_x,g_x}(x) &= \sum_{i=1}^n B_{h_x(i)}(x) E^* \left(\xi(T^*_{(i)}, \delta^*_{(i)}, x)\right) - \sum_{i=1}^n B_{g_x(i)}(x) \xi(T_{(i)}, \delta_{(i)}, x), \\ b_{h_x}(x) &= \sum_{i=1}^n B_{h_x(i)}(x) E \left(\xi(T_{(i)}, \delta_{(i)}, x)\right), \\ b(x) &= C^{5/2} \frac{1}{2} d_K (1 - p(x)) \mu(x)\end{aligned}$$

and $\sup_{x \in I} |r^*_n(x)| = O_{P^*}\left((nh_x)^{-3/4} (\ln n)^{3/4}\right)$ a.s.

Note that the main effect of the pilot bandwidth g_x in $MSE^*_{x,g_x}(h_x)$ is confined into the second term in (12), that can be decomposed as follows (see Lemma 5.1 in Van Keilegom and Veraverbeke, 1997a):

$$\begin{aligned}&(nh_x)^{1/2} |\hat{b}_{h_x,g_x}(x) - b_{h_x}(x)| \\ &= (nh_x)^{1/2}(1 - p(x)) \Bigg| \int_0^\infty \left(\text{bias}^* \left(\hat{H}^*_{h_x,g_x}(t|x)\right) - \text{bias}\left(\hat{H}_{h_x}(t|x)\right)\right) \frac{dH^1(t|x)}{(1 - H(t|x))^2} \\ &+ \left(\text{bias}^* \left(\hat{H}^{1*}_{h_x,g_x}(t|x)\right) - \text{bias}\left(\hat{H}^1_{h_x}(t|x)\right)\right) \frac{1}{1 - H(t|x)} \\ &- \int_0^\infty \left(\text{bias}^* \left(\hat{H}^{1*}_{h_x,g_x}(t|x)\right) - \text{bias}\left(\hat{H}^1_{h_x}(t|x)\right)\right) \frac{dH(t|x)}{(1 - H(t|x))^2} \Bigg|,\end{aligned}$$

where

$$\hat{H}^*_{h_x,g_x}(t|x) = 1 - \left(1 - \hat{F}^*_{h_x,g_x}(t|x)\right)\left(1 - \hat{G}^*_{h_x,g_x}(t|x)\right),$$
$$\text{bias}^*\left(\hat{H}^*_{h_x,g_x}(t|x)\right) = E^*\left(\hat{H}^*_{h_x,g_x}(t|x)\right) - \hat{H}_{g_x}(t|x),$$
$$\text{bias}\left(\hat{H}_{h_x}(t|x)\right) = E\left(\hat{H}_{h_x}(t|x)\right) - H(t|x)$$

and $\hat{G}^*_{h_x,g_x}(t|x)$ is the bootstrap version of the Beran estimator, $\hat{G}_h(t|x)$, defined as in (2) but replacing $\delta_{(i)}$ by $1 - \delta_{(i)}$. As in the proof of Lemma 4.1 in Van Keilegom and Veraverbeke (1997a), we have that

$$(nh_x)^{1/2} \sup_{t\in[a,b]} \left|\text{bias}\left(\hat{H}^*_{h_x,g_x}(t|x)\right) - \text{bias}\left(\hat{H}_{h_x}(t|x)\right)\right| \leq \frac{1}{2} d_K C^{5/2} \sup_{t\in[a,b]} |\hat{\ddot{H}}_{g_x}(t|x) - \ddot{H}(t|x)| + o(1)$$

where $\hat{\ddot{H}}_{g_x}(t|x) = \sum_{i=1}^n B^{(2)}_{g_x(i)}(x) I(T_{(i)} \leq t)$ is the kernel estimator of $\ddot{H}(t|x) = (\partial^2/\partial x^2) H(t|x)$ with NW weights and bandwidth g_x and

$$B^{(2)}_{g_x(i)}(x) = \frac{\partial^2}{\partial x^2} B_{g_x(i)}(x) = \frac{\partial^2}{\partial x^2}\left(\frac{K\left(\frac{x - X_{(i)}}{g_x}\right)}{\sum_{j=1}^n K\left(\frac{x - X_j}{g_x}\right)}\right).$$

Therefore, the pilot bandwidth g_x must be chosen to minimize the MISE of $\hat{\ddot{H}}_{g_x}(t|x)$.

Lemma 4 *Assume (A1)-(A4), (A6), (A8), (A9), $h_x = Cn^{-1/5}$ for some $C > 0$, $g_x \to 0$ and $ng_x^5 (\ln n)^{-1} \to \infty$. Then, the asymptotically optimal pilot bandwidth g_x for the bootstrap resampling, which minimizes the MISE of $\hat{\ddot{H}}_{g_x}(t|x)$, is:*

$$g_x = \left(\frac{5 c_{K''} m(x) \int_0^\infty H(t|x)(1 - H(t|x))\, dt}{\int_0^\infty \left(\frac{\partial^4}{\partial x^4}(H(t|x) m(x)) - \frac{\partial^2}{\partial x^2}(H(t|x) m''(x))\right)^2 dt}\right)^{1/9} n^{-1/9}. \tag{13}$$

The proof of Lemma 4 derives directly from the proof of the main result in Cao (1991). Because of this reason it is not included in the Appendix.

Remark: The bandwidth sequence $g_x = g_n$ is typically asymptotically larger than $h_x = h_n$. This oversmoothing pilot bandwidth is required for the bootstrap bias and variance to be asymptotically efficient estimators for the bias and variance terms. The order $n^{-1/9}$ of this asymptotically optimal pilot bandwidth satisfies the conditions in Theorem 1 of Li and Datta (2001), and is also equal to the order obtained by Cao and González-Manteiga (1993) for the uncensored case.

4 Simulation study

In this section we compare the proposed nonparametric estimators with the semiparametric estimators in Peng and Dear (2000), which are implemented in the *smcure* package

in R (Cai et al, 2012). These estimators assume a logistic expression for the incidence and a proportional hazards (PH) model for the latency.

We carry out a simulation study with two purposes. First, we evaluate the finite sample performance of the nonparametric estimators $1-\hat{p}_{h_x}$ and $\hat{S}_{0,h_x}$, both computed in a grid of bandwidths with the Epanechnikov kernel defined on $[-1,1]$, and we compare the results with those of the semiparametric estimators. Second, the practical behavior of the bootstrap bandwidth selector is assessed. We consider two different models and for both, the censoring times are generated according to the exponential distribution with mean $1/0.3$ and the covariate X is $U(-20,20)$.

Model 1: For comparison reasons, this simulated configuration is the same as the so-called mixture cure (MC) model considered in Xu and Peng (2014). The data are generated from a logistic-exponential MC model, where the probability of not being cured is

$$p(x)=\frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)},$$

with $\beta_0=0.476$ and $\beta_1=0.358$, and the survival function of the uncured subjects is:

$$S_0(t|x)=\begin{cases}\dfrac{\exp(-\lambda(x)t)-\exp(-\lambda(x)\tau_0)}{1-\exp(-\lambda(x)\tau_0)} & \text{if } t\leq\tau_0\\ 0 & \text{if } t>\tau_0\end{cases},$$

where $\tau_0=4.605$ and $\lambda(x)=\exp((x+20)/40)$. The percentage of censored data is 62% and of cured data is 53%. In Figure 1 we show the shape of the theoretical incidence and latency. In this model the incidence is a logistic function and the latency fulfills the PH assumption, so this model satisfies the assumptions of the semiparametric estimator by Peng and Dear (2000) and therefore it is expected to give very good results.

Model 2: The data are generated from a cubic logistic-exponential mixture model, where the probability of not being cured is:

$$p(x)=\frac{\exp(\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 x^3)}{1+\exp(\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 x^3)},$$

with $\beta_0=0.0476$, $\beta_1=-0.2558$, $\beta_2=-0.0027$ and $\beta_3=0.0020$, and the survival function of the uncured subjects is:

$$S_0(t|x)=\frac{1}{2}\left(\exp(-\alpha(x)t^5)+\exp(-100t^5)\right),$$

with

$$\alpha(x)=\frac{1}{5}\exp((x+20)/40).$$

The percentages of censored and cured data are 58% and 47%, respectively. Figure 1 gives the shape of the theoretical incidence and latency in this model.

The incidence is not a logistic function and the effect of the covariate on the failure time of the uncured patients does not fit a PH model. So, the results will show the gain of using the proposed nonparametric estimators, that do not require any parametric or semiparametric assumptions, with respect to the semiparametric ones.

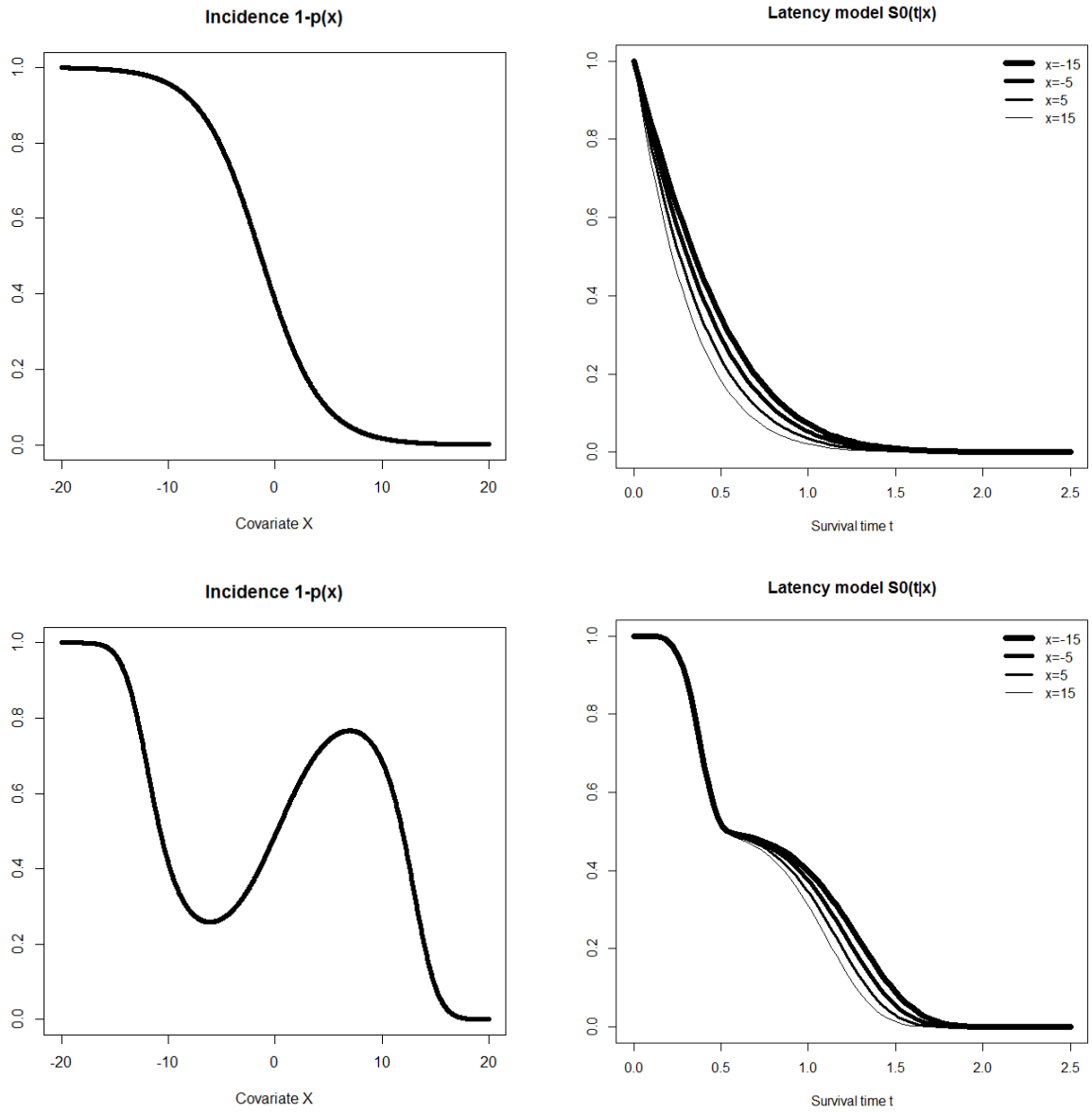

Fig. 1. Theoretical incidence (left) and latency (right) in Model 1 (top) and Model 2 (bottom).

4.1 Efficiency of the nonparametric estimators

A total of $m = 1000$ samples of size $n = 100$ are drawn to approximate, by Monte Carlo, the mean squared error (MSE) of the incidence estimators, and the mean integrated squared error (MISE) of the latency estimators, for a grid of bandwidths from $h_0 = 1.2$ to $h_{99} = 20$ for the incidence function, and from $h_0 = 10$ to $h_{99} = 40$ for the latency. The results for both models are shown in Figure 2.

Regarding the MSE of the incidence estimators in Model 1, Figure 2 shows that there is a range of bandwidths, from $h = 4.8$ to $h = 8.5$ (light blue lines) for which the nonparametric estimator is quite competitive with respect to the semiparametric estimator in some values x of the covariate, and it works much better when the value of the covariate is around 0. In Model 2, as expected, the nonparametric estimator outperforms the semiparametric one, regardless how small or large the bandwidth is, except for the extreme values of the interval $[-20, 20]$.

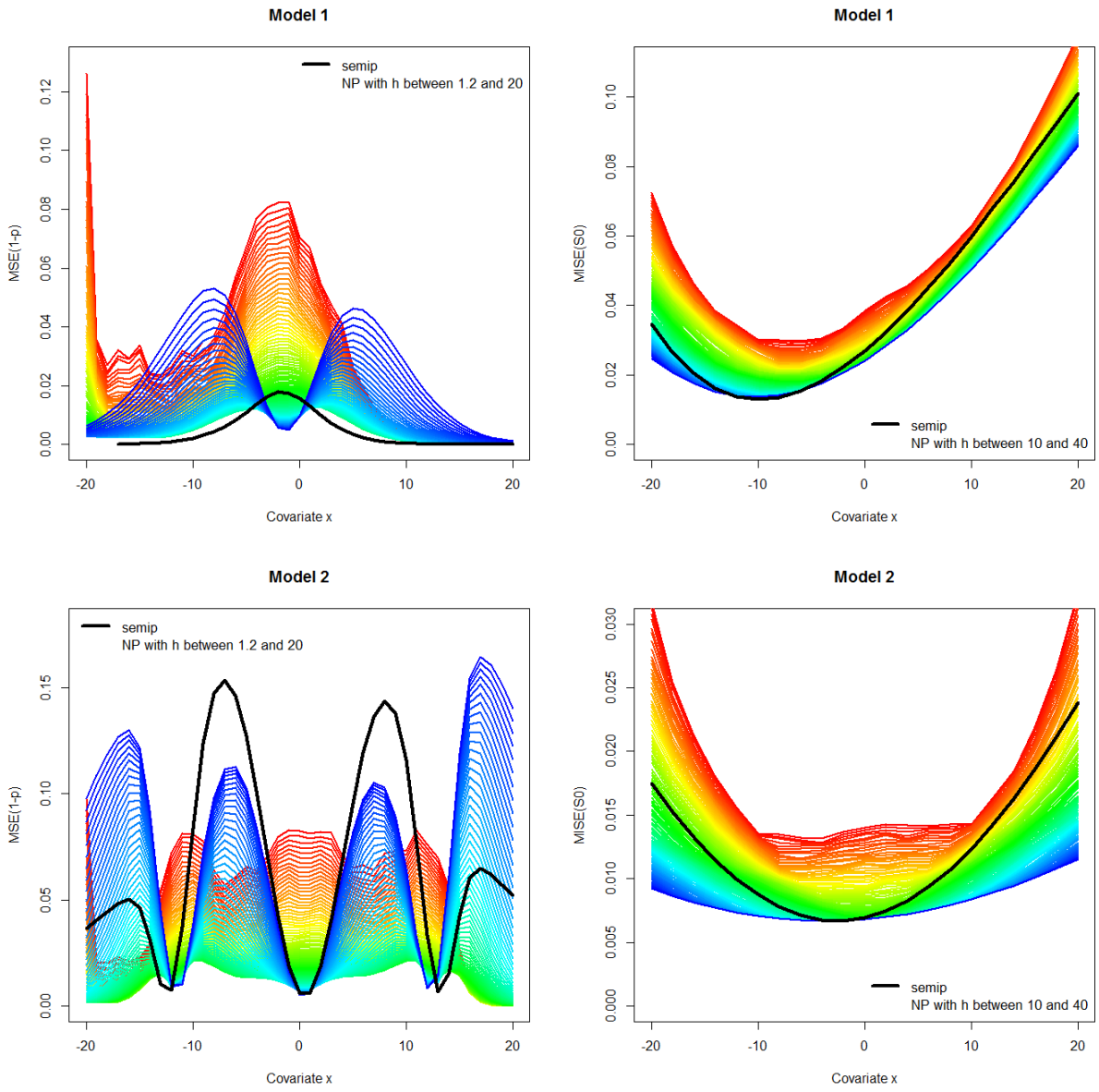

Fig. 2. On the left, MSE for the semiparametric (black line) and the nonparametric estimators of $1 - p(x)$ computed with different bandwidths: from $h_0 = 1.2$ (red line) to $h_{99} = 20$ (blue line). On the right, MISE for the semiparametric (black line) and the nonparametric estimators of $S_0(t|x)$ computed with different bandwidths: from $h_0 = 10$ (red line) to $h_{99} = 40$ (blue line). The data were generated from Model 1 (top) and Model 2 (bottom).

Considering the latency estimators, it is noteworthy that in Model 1, for values of the covariate greater than 0, there is a wide range of bandwidths for which the MISE of the nonparametric estimator is smaller than the MISE of the semiparametric estimator, as we can see in Figure 2 (top, right). In Model 2, the nonparametric estimator of the latency outperforms the semiparametric estimator for all the values of the covariate with almost any value of the bandwidth, regardless how large it is, as long as the bandwidth is greater than 17.

In short, the nonparametric estimators are quite comparable to the semiparametric ones in situations where the assumptions of the semiparametric estimator are fulfilled, and they outperform the semiparametric estimators when the incidence is not a logistic function and the latency does not fit a PH model. The efficiency of the nonparametric estimators depends on the choice of the bandwidth, but although the optimal value of the bandwidth

remains unknown, the simulations show that, for quite wide ranges of bandwidth values, the proposed nonparametric methods outperform the existing semiparametric estimator of Peng and Dear (2000).

4.2 Efficiency of the bootstrap bandwidth selector

In this simulation study, we consider sample sizes of $n = 50, 100$ and 200. For a number of $m = 1000$ trials, we approximate the MSE_x of the proposed nonparametric estimator of the incidence, evaluated in the optimal bandwidth $h_{x,MSE}$. The bootstrap $MSE^*_{x,g_x}(h_x)$ evaluated in the bootstrap bandwidth h^*_{x,g_x} is also computed.

Note that in order to minimize MSE^*_{x,g_x} in h_x for each value x of the covariate, since it is a computationally expensive algorithm, we carry out a two-step method with a double search in each stage. In the first step, we draw $B = 80$ bootstrap resamples and consider a number of 21 bandwidths equispaced on a logarithmic scale, from $h_0 = 0.2$ to $h_{20} = 25$ in the first search, whereas in the second search the grid is centered around the optimal bandwidth obtained in the first search. Then, we carry out the second step with also a double search in a similar way we did for the first step, but now with two differences: we draw $B = 1000$ bootstrap resamples and we consider a finer smaller grid of 5 bandwidths in both the first and second search.

The asymptotically optimal pilot bandwidth g_x has a quite involved expression, see (13), which depends on unknown functions in a rather awkward way. In view of the fact that the choice of g_x has a low effect on the final bootstrap bandwidth, we propose to use a naive selector, keeping the $n^{-1/9}$ optimal order. Since the distribution of the covariate is uniform, we consider the following global pilot bandwidth, that does not depend on the value x for which the estimation is to be carried out:

$$g = \frac{X_{(n)} - X_{(1)}}{10^{7/9}} n^{-1/9}. \tag{14}$$

Note that, on the account of $X \in U[-20, 20]$, when $n = 100$ the value of the global pilot bandwidth g is $(X_{(n)} - X_{(1)})/10 \simeq 4$. Similarly, $g \simeq 4.32$ ($g \simeq 3.70$) when $n = 50$ ($n = 200$). For a naive pilot bandwidth selector if the distribution for X can not be assumed uniform, see Section 5.

Figure 3 shows the mean and the 25th and 75th percentiles of MSE^*_{x,g_x} evaluated at the proposed bootstrap bandwidth, along the $m = 1000$ simulated samples. The value of the MSE_x of the nonparametric estimator, approximated by Monte Carlo and evaluated at the MSE bandwidth $h_{x,MSE}$, is also given as reference. We can observe that the mean and the 25th and 75th percentiles of MSE^*_{x,g_x} approach MSE_x properly. As expected, the similarity increases with the sample size. Moreover, we can also check how MSE_x and MSE^*_{x,g_x} decrease as n becomes larger.

The performance of the bootstrap bandwidth for Models 1 and 2 is shown in Figure 4. The optimal $h_{x,MSE}$, approximated by Monte Carlo, is displayed together with the mean and the 25th and 75th percentiles of the 1000 bootstrap bandwidths h^*_x. We can appreciate how the bootstrap bandwidth h^*_x approaches $h_{x,MSE}$, adapting properly to the shape of $h_{x,MSE}$ for any sample size. The optimal bandwidth $h_{x,MSE}$ has got peaks at the values x of the covariate for which $p''(x) = 0$. In other terms, those peaks only occur at points x

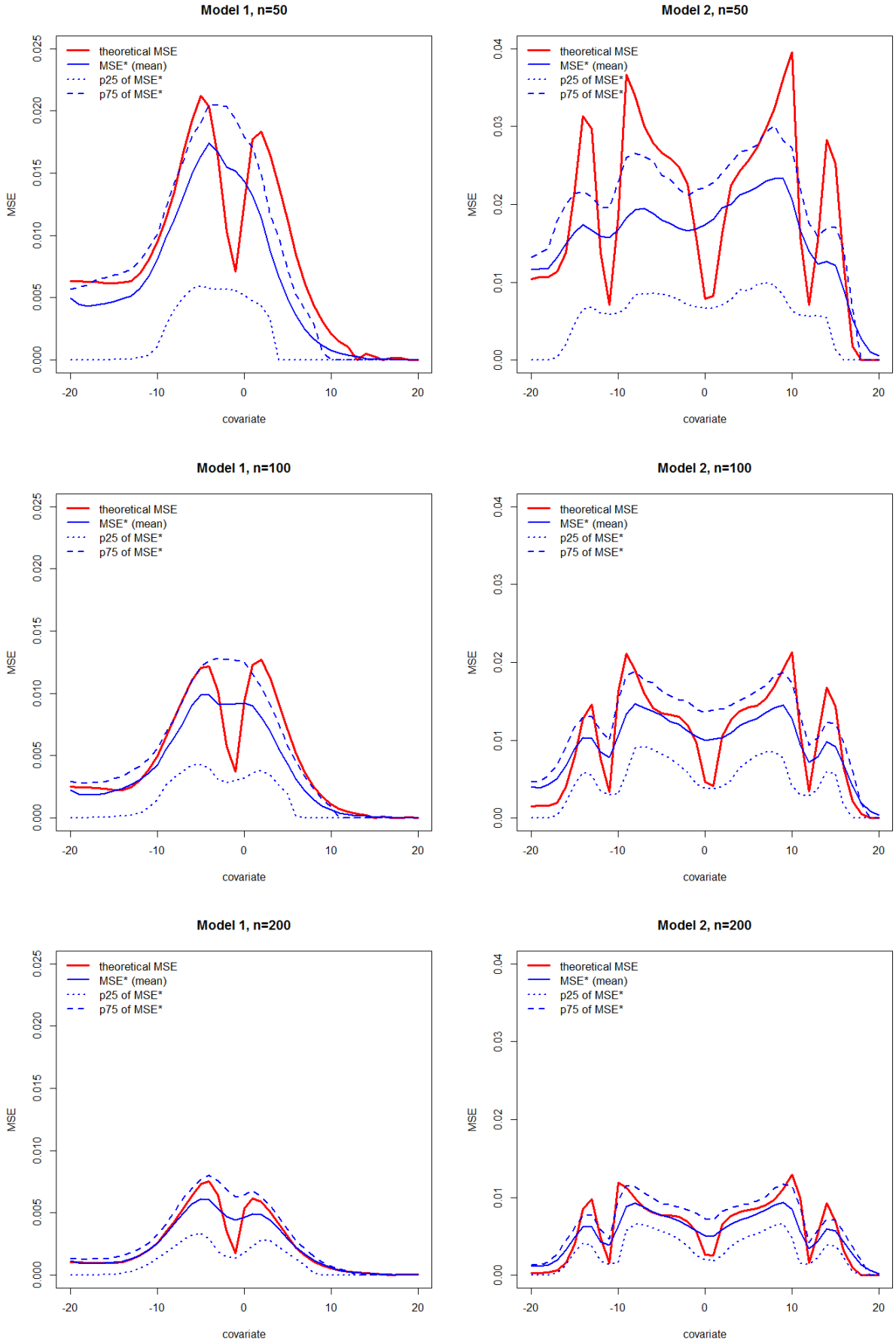

Fig. 3. Optimal MSE_x of the nonparametric estimator of the incidence (red line) evaluated at the optimal MSE bandwidth, and mean (blue line), 25th (dotted blue line) and 75th (dashed blue line) percentiles of $MSE^*_{x,g_x}(h_x)$ computed with the bootstrap bandwidth along $m = 1000$ samples, with sample sizes $n = 50$ (top), $n = 100$ (center) and $n = 200$ (bottom), for Model 1 (left) and Model 2 (right).

for which the optimal bandwidth is infinitely large because the best choice is to smooth as much as possible, and the best local fit is a global fit. Note that if such large bandwidths are used, those values of x correspond to the values where the MSE_x shows deep valleys, that is, there is a noticeable improvement in the estimation of the incidence.

5 Application to real data

We applied both the semiparametric and the nonparametric estimators to a real dataset of 414 colorectal cancer patients from CHUAC (Complejo Hospitalario Universitario de A Coruña), Spain. We considered two covariates: the stage, from 1 to 4, and the age, from 23 to 103. About 50% of the observations are censored, with the percentage of censoring varying from 30% to almost 71%, depending on the stage. In Table 1 we show a summary of the data set.

Table 1: Colorectal cancer patients from CHUAC

Stage	Number of patients	Number of censored data	% Censoring
1	62	44	70.97
2	167	92	55.09
3	133	53	39.85
4	52	16	30.77
	414	205	49.52

The incidence in the four different stages of the disease is computed with both the semiparametric and the nonparametric estimators. The age of the patients has been considered as the covariate.

For the nonparametric estimator of $1-p$, a naive pilot bandwidth selector has been proposed in (14) if the distribution of X is uniform. The idea is to provide a data-driven pilot bandwidth which only depends on both the sample size and on the distribution of the covariate, keeping the $n^{-1/9}$ optimal order. Taking into account that in this case the distribution of the covariate is not uniform (see Figure 5), we propose to use the following local pilot bandwidth:

$$g_x = \frac{d_k^+(x) + d_k^-(x)}{2} 100^{1/9} n^{-1/9}, \tag{15}$$

where $d_k^+(x)$ is the distance from x to the k-th nearest neighbor on the right, $d_k^-(x)$ the distance from x to the k-th nearest neighbor on the left, and k a suitable integer depending on the sample size. If there are not at least k neighbors on the right (or left), we use $d_k^+(x) = d_k^-(x)$ (or $d_k^-(x) = d_k^+(x)$) respectively. Our numerical experience shows that a good choice is to consider $k = n/4$. Note that when $n = 100$ the value of the local pilot bandwidth g_x is the mean distance to the 25th nearest neighbor on both the left and right sides.

Figure 5 shows the estimations of the probability of being cured for the different stages with respect to the age of the patients. For the nonparametric estimator, alongside the

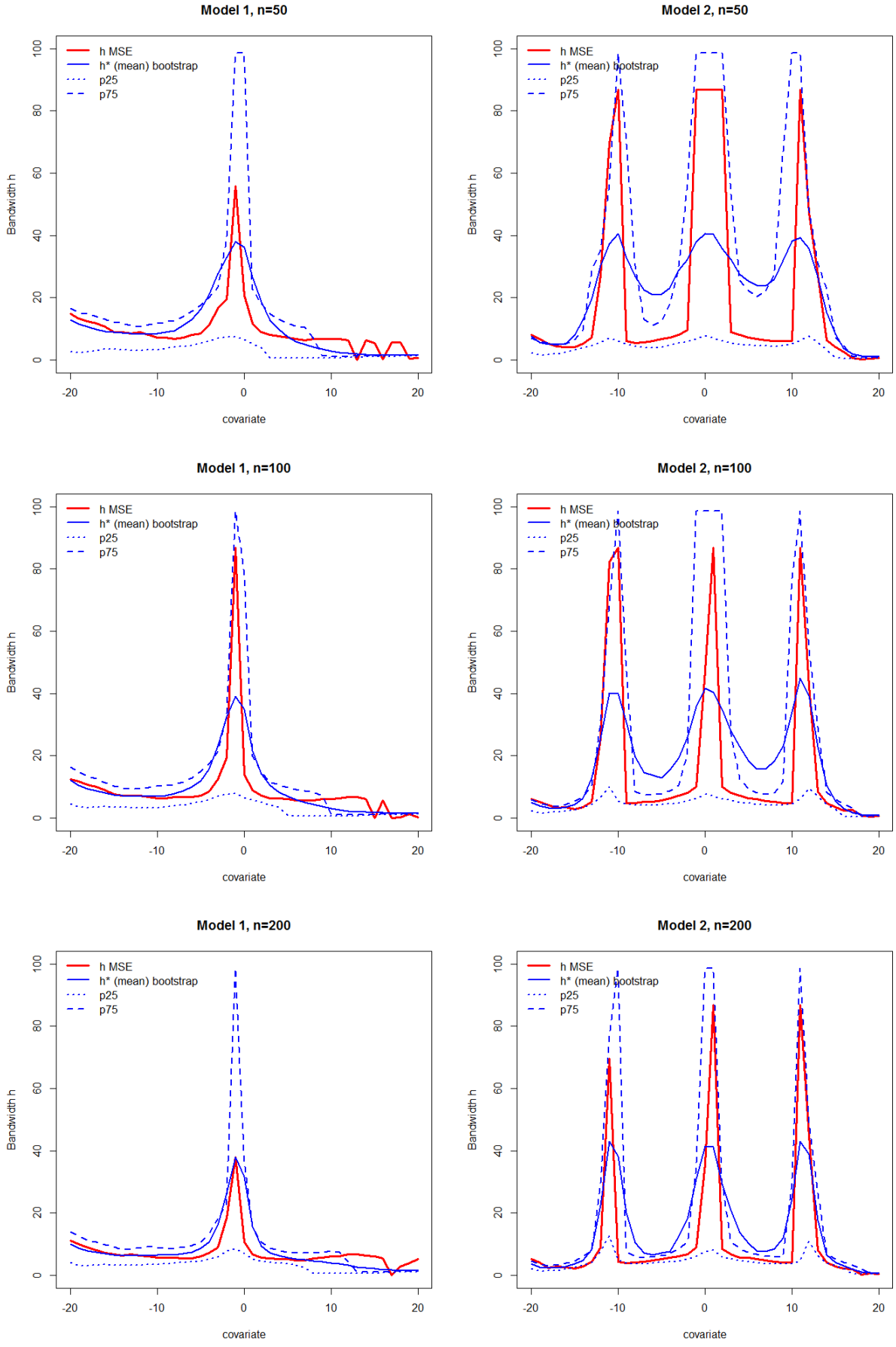

Fig. 4. Optimal $h_{x,MSE}$ (red line), mean (blue line) and the 25th (dotted blue line) and 75th (dashed blue line) percentiles of the bootstrap bandwidth h_x^* along $m = 1000$ samples, with sample sizes $n = 50$ (top), $n = 100$ (center) and $n = 200$ (bottom), for Model 1 (left) and Model 2 (right).

bootstrap bandwidth, we have also used a smoothed bandwidth, considering 5 neighbors on each side. We followed Cao, Janssen and Veraverbeke (2001), who applied a method for smoothing local bandwidths in another context. We can see that the effect of the age on the probability of being cured changes with the stage. For example, in Stage 1, patients have a probability of survival between 25% and 65%, depending on the age; whereas in Stage 3, for patients above 60, in a 10 years gap that probability decreases considerably from 40% to almost 0%. It is important to highlight the difference between the nonparametric and the semiparametric curves, that seems to indicate that the logistic model is not valid for the data. The results in Stage 4 deserve some comments. A total of 11 in the 12 greatest lifetimes in Stage 4, including the largest lifetime, are uncensored and, consequently, uncured. This causes that the nonparametric estimation of the probability of being cured is equal to 0. Although it should not be stated that it is impossible for a patient with Stage 4 colorectal cancer to survive, this estimation reinforce the assertion that long-term survival in patients with Stage 4 colorectal cancer is uncommon (Miyamoto et al, 2015). This fact, far from being a weakness of the nonparametric method, is an important advantage, since it allows to detect situations in which introducing the possibility of cure does not contribute to improve the model.

Note that in order to obtain the optimal bootstrap bandwidth, $B = 1000$ bootstrap resamples are used. In a similar way as we did in Section 4.2, we carry out a one-step procedure with a double search. We consider a number of 21 bandwidths equispaced on a logarithmic scale in both searches. The first search is performed between 0.2 and the empirical range of X. The second one is carried out using another grid centered around the optimal bandwidth obtained in the first search. We show the resulting bootstrap bandwidths, with the corresponding local pilot bandwidths, for the different values of the covariate age in Figure 6.

In Figures 7 and 8 we show the latency estimation for Stages $1, 2, 3$ and 4 for two different ages, 45 and 76. The nonparametric estimator $\hat{S}_{0,h_x}$ is computed with five different constant bandwidths: $h = 10, 15, 20, 25$ and 30. It is noteworthy that in Stages 1 and 2 for 45 years, the bandwidth selection influences considerably in the latency estimation. This is due to the low density of the covariate around this age, as we can see in Figure 5.

6 Acknowledgements

The first author's research was sponsored by the Spanish FPU grant from MECD with reference FPU13/01371. The work of the first author has been partially carried out during a visit at the Université catholique de Louvain, financed by INDITEX. All the authors acknowledge partial support by the MINECO grant MTM2014-52876-R (EU ERDF support included). The first three authors' research has been partially supported by MICINN Grant MTM2011-22392 (EU ERDF support included) and Xunta de Galicia GRC Grant CN2012/130. The research of the fourth author was supported by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy), and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain").

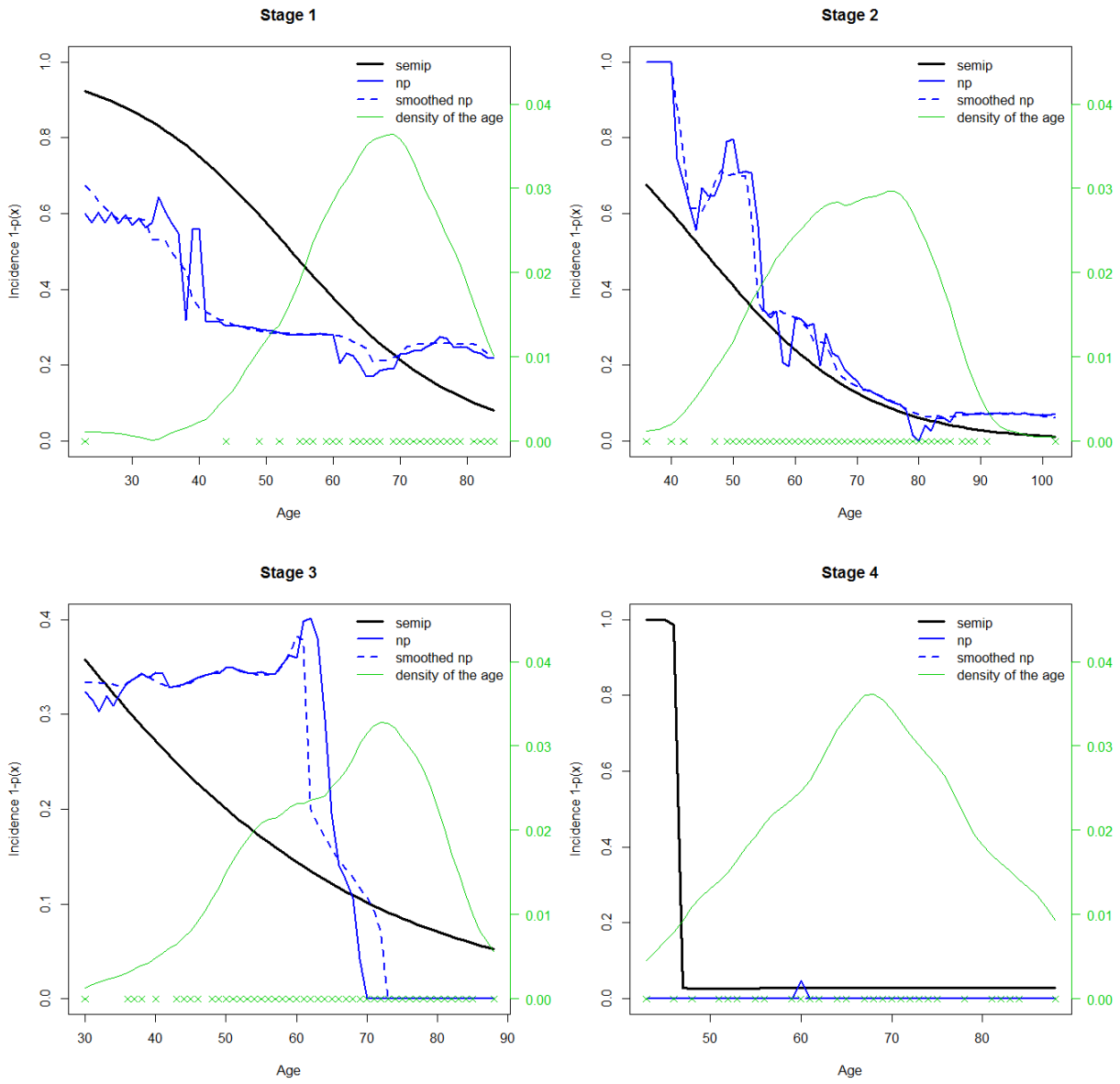

Fig. 5. Semiparametric (black line) and nonparametric estimators of the probability of cure (incidence) of the patients in Stages 1-4 depending on the age computed with the bootstrap bandwidth h_x^* (blue line) and with the smoothed h_x^* (dashed blue line). The green line represents the Parzen-Rosenblatt density estimator of the covariate age, using a plug-in bandwidth.

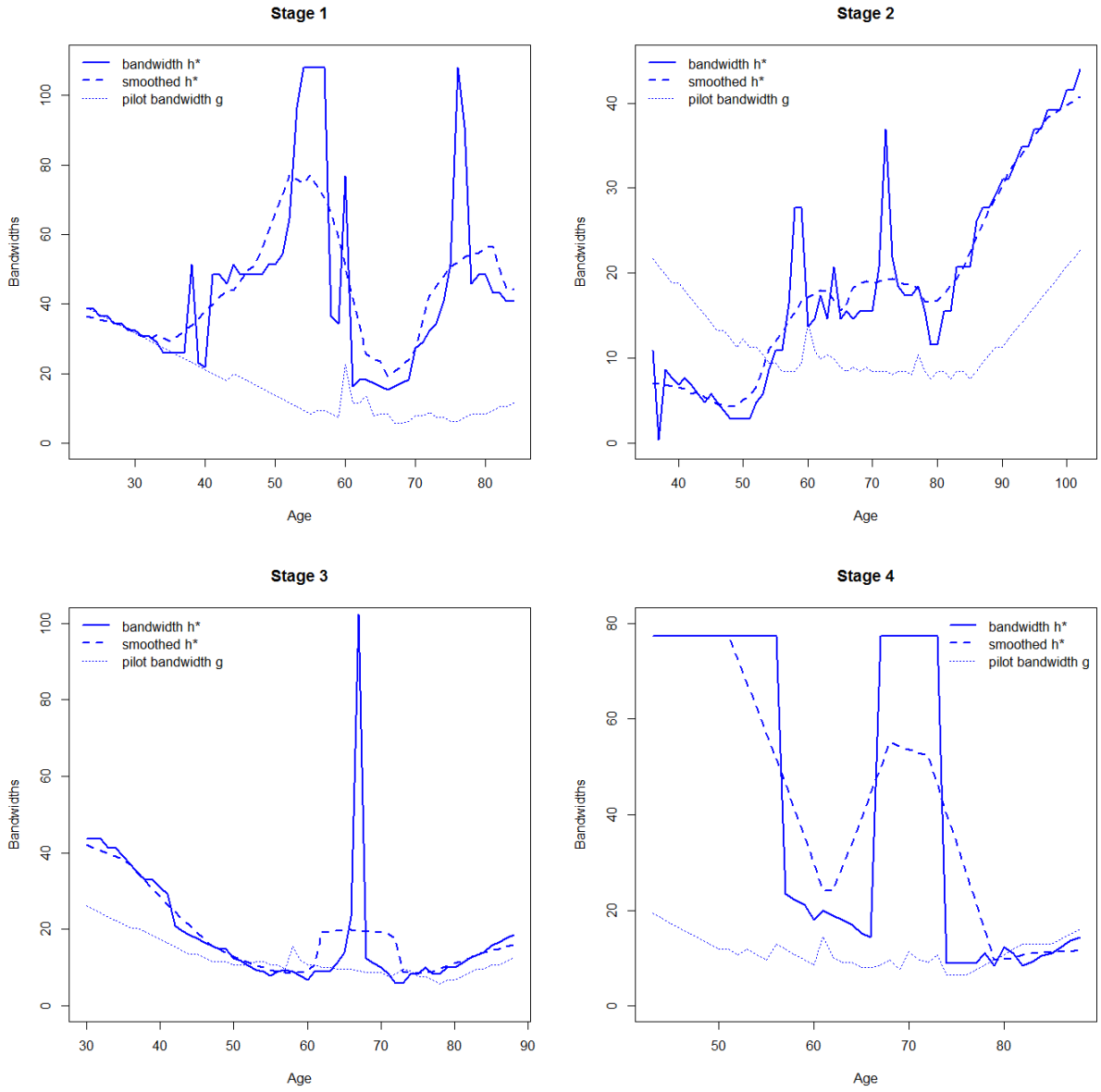

Fig. 6. Bootstrap bandwidth (blue line), smoothed bootstrap bandwidth (dashed blue line) and local pilot bandwidth g_x (dotted blue line) used for the nonparametric incidence estimator for patients in Stages 1-4.

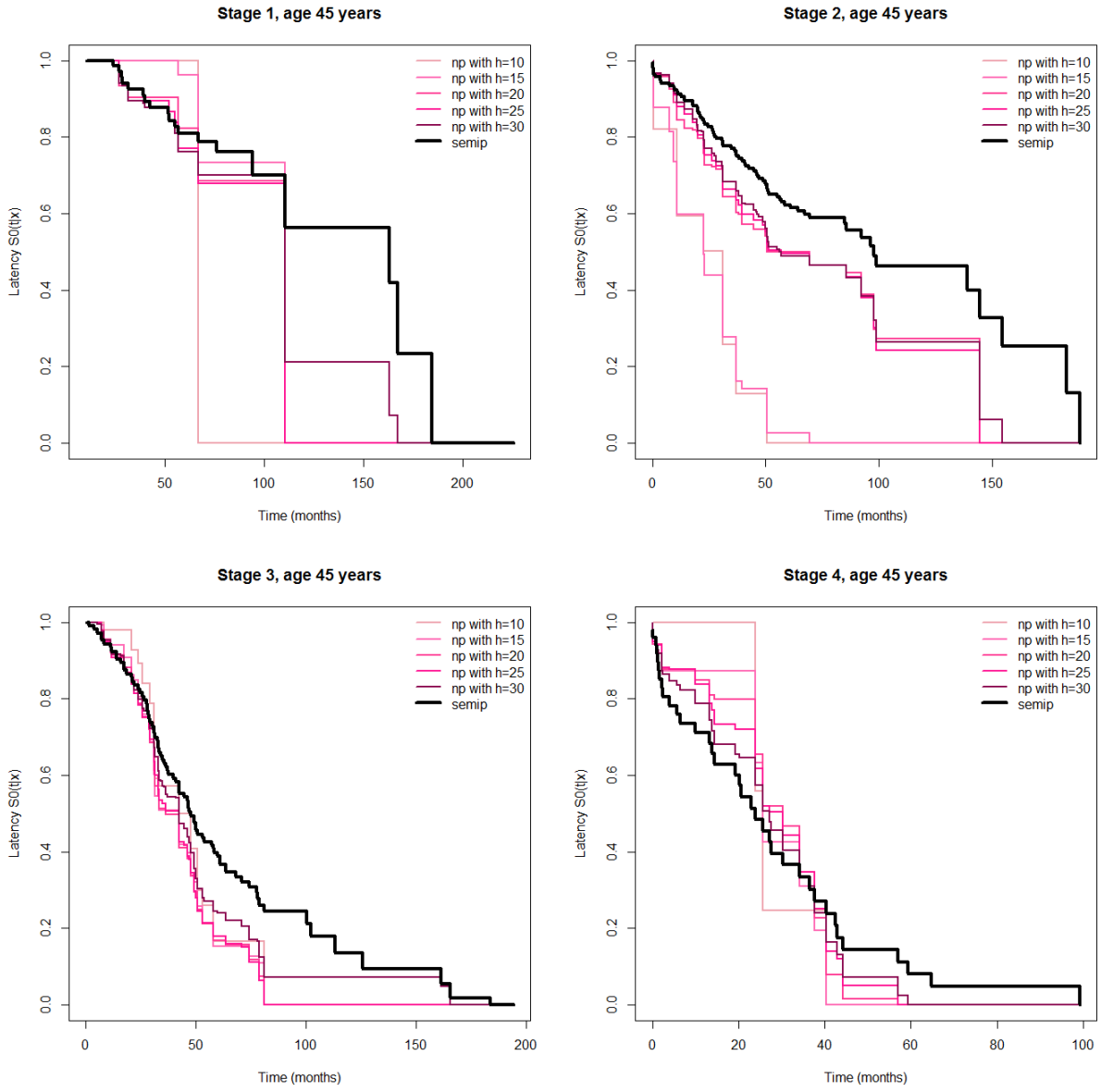

Fig. 7. Estimated latency for patients of age 45 in Stages 1-4, using the semiparametric (black line) and nonparametric estimators with 5 equispaced bandwidths ranging from $h_0 = 10$ (light pink line) to $h_4 = 30$ (dark pink line).

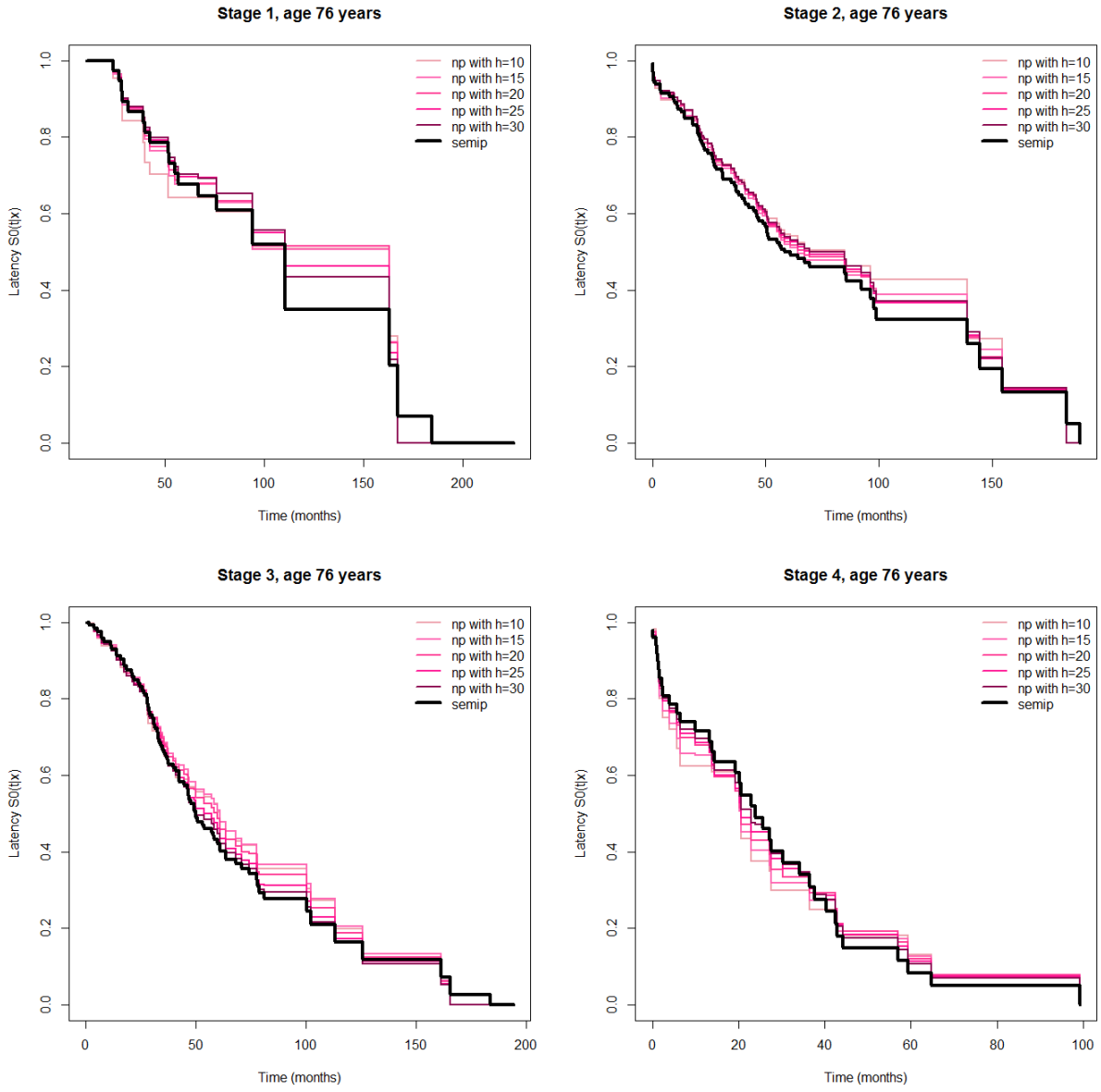

Fig. 8. Estimated latency for patients of age 76 in Stages 1-4, using the semiparametric (black line) and nonparametric estimators with 5 equispaced bandwidths ranging from $h_0 = 10$ (light pink line) to $h_4 = 30$ (dark pink line).

Appendix

Proof of Lemma 1. Suppose we have two formulations of model (1):

$$S(t|x) = 1 - p(x) + p(x)S_0(t|x) \text{ and } S^*(t|x) = 1 - p^*(x) + p^*(x)S_0^*(t|x).$$

We need to show that $S(t|x) = S^*(t|x)$ if and only if $p(x) = p^*(x)$ and $S_0(t|x) = S_0^*(t|x)$ for all $x \in D$ and $t < T^+$. The "if"part is clearly true in all cases, so we concentrate on "only if": suppose that $S(t|x) = S^*(t|x)$, then, rearranging Eq. (1) gives the ratio:

$$\frac{p\,(x)}{p^*\,(x)} = \frac{1 - S_0^*(t|x)}{1 - S_0(t|x)} = c\,(x) \quad \text{for all } t < T^+ \text{ and for all } x \in D. \tag{A.1}$$

In particular,

$$S_0^*(t|x) = 1 - c\,(x)\,(1 - S_0(t|x)) \quad \text{for all } t < T^+ \text{ and for all } x \in D.$$

For a time T^+ large enough such that $S_0^*(T^+|x) = S_0(T^+|x) = 0$, we have

$$0 = S_0^*(T^+|x) = 1 - c\,(x)\left(1 - S_0(T^+|x)\right) = 1 - c\,(x) \text{ for all } x \in D.$$

Hence, $c(x)$ is constant and equal to one for all x and thus, from (A.1), $p(x) = p^*(x)$ and $S_0(t|x) = S_0^*(t|x)$, so $S(t|x)$ is uniquely represented by $1 - p\,(x) + p\,(x)\,S_0(t|x)$. □

Proof of Theorem 2. The idea is to estimate $p\,(x)$ locally, maximizing the observed local likelihood function around x. It can be proved that the maximum likelihood estimator of the survival function $S_0(t|x) = 1 - F_0(t|x)$ has jumps only at the observations $(X_i, T_i, \delta_i)\,, i = 1, \ldots, n$ with jumps

$$q_i\,(x) = S_0(T_{(i)}^-|x) - S_0(T_{(i)}|x).$$

The local likelihood of the model is

$$L\left(p\,(x)\,, S_0\,(\cdot|x)\right) = \prod_{i=1}^{n}\left\{\left[p\,(x)\,q_i\,(x)\right]^{B_{h(i)}(x)\delta_{(i)}}\left[1 - p\,(x) + p\,(x)\left(1 - \sum_{j=1}^{i-1} q_j\,(x)\right)\right]^{\left(1-\delta_{(i)}\right)B_{h(i)}(x)}\right\}.$$

Let $D_i\,(x) = B_{h(i)}\,(x)\,\delta_{(i)}$ and $P_i\,(x) = p\,(x)\,q_i\,(x)$, then

$$L\left(p\,(x)\,, S_0\,(\cdot|x)\right) = \prod_{i=1}^{n}\left\{P_i\,(x)^{D_i(x)}\left(1 - \sum_{j=1}^{i-1} P_j\,(x)\right)^{B_{h(i)}(x) - D_i(x)}\right\}.$$

Consider now the functions $\lambda_i\,(x) = P_i\,(x)\,/\left(1 - \sum_{j=1}^{i-1} P_j\,(x)\right)$ satisfying

$$1 - \sum_{j=1}^{k} P_j\,(x) = \prod_{j=1}^{k}(1 - \lambda_j\,(x)). \tag{A.2}$$

Straightforward calculations yield

$$L\left(\lambda_1(x),\ldots,\lambda_n(x)\right)=\prod_{i=1}^{n}\lambda_i\left(x\right)^{D_i(x)}\left(1-\lambda_i\left(x\right)\right)^{\sum_{r=i+1}^{n}B_{h(r)}(x)}.$$

Maximizing the likelihood of the observations for the cure model is equivalent to maximizing

$$\max_{\lambda_i\geq 0;i=1,\ldots,n}\Psi(\lambda_1,\ldots,\lambda_n),$$

where Ψ is the local loglikelihood:

$$\Psi(\lambda_1\left(x\right),\ldots,\lambda_n\left(x\right))=\sum_{i=1}^{n}\left[D_i\left(x\right)\log\lambda_i\left(x\right)+\left(\sum_{r=i+1}^{n}B_{h(r)}\left(x\right)\right)\log\left(1-\lambda_i\left(x\right)\right)\right]$$

subject to

$$\prod_{i=1}^{n}(1-\lambda_i\left(x\right))=1-\sum_{j=1}^{n}P_j\left(x\right)=1-p\left(x\right). \tag{A.3}$$

Using standard maximization techniques, we obtain

$$\widehat{\lambda}_i\left(x\right)=\frac{D_i\left(x\right)}{\sum_{r=i+1}^{n}B_{h(r)}\left(x\right)+D_i\left(x\right)}=\frac{\delta_{(i)}B_{h(i)}\left(x\right)}{\sum_{r=i+1}^{n}B_{h(r)}\left(x\right)+\delta_{(i)}B_{h(i)}\left(x\right)}.$$

Replacing λ_i in (A.3) by $\widehat{\lambda}_i(x)$, we obtain the estimator of $1-p\left(x\right)$ given in (4).

With respect to the distribution of the uncured subjects, note that

$$F_0\left(T_{(i)}|x\right)=\sum_{j=1}^{i}q_j\left(x\right).$$

Since the jumps satisfy $P_i\left(x\right)=p\left(x\right)q_i\left(x\right)$ and using (A.2), we find that the local maximum likelihood estimator is given by

$$\widehat{F}_0\left(T_{(i)}|x\right)=\frac{1}{\widehat{p}_h\left(x\right)}\left[1-\prod_{j=1}^{i}\left(1-\widehat{\lambda}_j\left(x\right)\right)\right]=\frac{\widehat{F_h}\left(T_{(i)}|x\right)}{\widehat{p}_h\left(x\right)},$$

with $\widehat{F_h}\left(T_{(i)}|x\right)$ the Beran estimator of $F=1-S$ computed at time $T_{(i)}$. □

The following auxiliary results are necessary to prove Theorem 3.

Lemma 5 (Xu and Peng (2014)) *Under assumption (A10),*

$$T^1_{\max}=\max_{i:\delta_i=1}\left(T_i\right)\to\tau_0 \text{ in probability as } n\to\infty.$$

Lemma 6 *Under assumption (A7), we have that*

$$n^{\alpha}(\tau_0-T^1_{\max})\to 0 \quad a.s.$$

for any $\alpha \in (0,1)$. In particular, for a sequence of bandwidths satisfying $nh^5(\ln n)^{-1} = O(1)$, we have

$$\tau_0 - T^1_{\max} = o\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \text{ a.s.} \tag{A.4}$$

Proof: Using the Borel-Cantelli lemma, it is sufficient to prove that

$$\sum_{n=1}^{\infty} P\left(|a_n(\tau_0 - T^1_{\max})| > \epsilon\right) < \infty, \text{ for all } \epsilon > 0, \tag{A.5}$$

where $a_n = n^\alpha$. Let us fix $\epsilon > 0$ and consider:

$$\begin{aligned}
P(|a_n(\tau_0 - T^1_{\max})| > \epsilon) =& P\left(T^1_{\max} < \tau_0 - \frac{\epsilon}{a_n}\right) \\
=& P\left(T_i < \tau_0 - \frac{\epsilon}{a_n}, \text{ for all } i = 1, 2, \ldots n \text{ where } \delta_i = 1\right) \\
=& E\left[P\left(T_i < \tau_0 - \frac{\epsilon}{a_n}, \text{ for all } i = 1, 2, \ldots n \text{ where } \delta_i = 1 \middle| \delta_1, \delta_2, \ldots, \delta_n\right)\right] \\
=& E\left[\prod_{i=1}^{n} P\left(T_i < \tau_0 - \frac{\epsilon}{a_n} \middle| \delta_i = 1\right)^{\delta_i}\right] = E\left[P\left(T_1 < \tau_0 - \frac{\epsilon}{a_n} \middle| \delta_1 = 1\right)^{\sum_{i=1}^n \delta_i}\right] \\
=& E\left[\left(H_{c,1}\left(\tau_0 - \frac{\epsilon}{a_n}\right)\right)^{\sum_{i=1}^n \delta_i}\right],
\end{aligned}$$

where

$$H_{c,1}(t) = P\left(T < t \middle| \delta = 1\right) = \frac{P(T < t, \delta = 1)}{P(\delta = 1)} = \frac{H_1(t)}{\rho},$$

with $\rho = P(\delta = 1) = E(\delta)$ and $H_1(t) = P(T < t, \delta = 1)$. Consequently, since $\sum_{i=1}^n \delta_i \overset{\text{d}}{=} Bi(n, \rho)$, we get:

$$\begin{aligned}
P(|a_n(\tau_0 - T^1_{\max})| > \epsilon) =& E\left[H_{c,1}\left(\tau_0 - \frac{\epsilon}{a_n}\right)^{\sum_{i=1}^n \delta_i}\right] = \sum_{j=0}^{n} \binom{n}{j} \rho^j (1-\rho)^{n-j} H_{c,1}\left(\tau_0 - \frac{\epsilon}{a_n}\right)^j \\
=& \sum_{j=0}^{n} \binom{n}{j} \left[\rho H_{c,1}\left(\tau_0 - \frac{\epsilon}{a_n}\right)\right]^j (1-\rho)^{nj} = \left[\rho H_{c,1}\left(\tau_0 - \frac{\epsilon}{a_n}\right) + 1 - \rho\right]^n \\
=& \left[\rho\left(H_{c,1}(\tau_0) - \frac{\epsilon}{a_n} H'_{c,1}(\tau_0) + \frac{\epsilon^2}{2a_n^2} H''_{c,1}(\xi_n)\right) + 1 - \rho\right]^n \\
=& \left[\rho - \rho\frac{\epsilon}{a_n} H'_{c,1}(\tau_0) + \rho\frac{\epsilon^2}{2a_n^2} H''_{c,1}(\xi_n) + 1 - \rho\right]^n \\
=& \left(1 - \rho\frac{\epsilon}{a_n} H'_{c,1}(\tau_0) + \rho\frac{\epsilon^2}{2a_n^2} H''_{c,1}(\xi_n)\right)^n,
\end{aligned} \tag{A.6}$$

for some $\xi_n \in \left[\tau_0 - \frac{\epsilon}{a_n}, \tau_0\right]$, since $H_{c,1}(\tau_0) = 1$.

Using assumption (A7), $\sup_{t \geq 0} |H''_{c,1}(t)| = C < \infty$. As a consequence, since $\epsilon/a_n \to 0$ as $n \to \infty$, then there exists some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$:

$$\left|\rho\frac{\epsilon^2}{2a_n^2} H''_{c,1}(\xi_n)\right| \leq \frac{\rho\epsilon^2}{2a_n^2} C \leq \rho\frac{\epsilon}{2a_n} H'_{c,1}(\tau_0). \tag{A.7}$$

From (A.6) and (A.7), we have that:

$$P(|a_n(\tau_0 - T^1_{\max})| > \epsilon) \leq \left(1 - \rho\frac{\epsilon}{2a_n}H'_{c,1}(\tau_0)\right)^n = \left(1 - \frac{\epsilon}{2a_n}H'_1(\tau_0)\right)^n = b_n^{n/a_n}, \quad \text{(A.8)}$$

where

$$b_n = \left(1 - \frac{\epsilon}{2a_n}H'_1(\tau_0)\right)^{a_n} \xrightarrow[n\to\infty]{} r, \quad \text{(A.9)}$$

with $r = \exp\left(-\frac{\epsilon H'_1(\tau_0)}{2}\right) < 1$.

Using (A.8) and (A.9), to prove (A.5) it suffices to show that $\sum_{n=1}^{\infty} r^{n/a_n} < \infty$. For that purpose, we will prove that

$$r^{n/a_n} < n^{-2}, \text{ for } n \text{ large enough} \quad \text{(A.10)}$$

and, since the hyperharmonic series $\sum_{n=1}^{\infty} n^{-2}$ is convergent, the series $\sum_{n=1}^{\infty} r^{n/a_n}$ will also be convergent.

Note that inequality (A.10) can be written as

$$2\log_R n < \frac{n}{a_n} \quad \text{(A.11)}$$

with $R = r^{-1} \in (1, \infty)$. Recall that $a_n = n^\alpha$ for some $\alpha \in (0, 1)$. Now condition (A.11) becomes

$$2\log_R n < n^{1-\alpha},$$

which is true for n large enough, since $n^{-(1-\alpha)}2\log_R n \to 0$. As a consequence, $n^\alpha(\tau_0 - T^1_{\max}) \to 0$ a.s. for any $\alpha \in (0, 1)$. On the other hand, note that:

$$\frac{n^{-\alpha}}{\left(\frac{\ln n}{nh}\right)^{3/4}} = \left[\frac{nh^5}{\ln n}\frac{n^{4-20\alpha/3}}{(\ln n)^4}\right]^{3/20} \xrightarrow[n\to\infty]{} 0$$

for $\alpha \geq 3/5$ and a sequence of bandwidths verifying $(\ln n)^{-1}nh^5 = O(1)$. Therefore, the result in (A.4) holds. This completes the proof. □

In the next three lemmas, we use existing results in the literature for a fixed t such that $1 - H(t|x) \geq \theta > 0$ in $(t, x) \in [a, b] \times I_\delta$, and apply them to the random value $t = T^1_{\max}$. Note that if $\tau_0 < \tau_G(x) = \tau_H(x)$ for all $x \in I_\delta$, then from Lemma 5 under assumption (A10), we have that:

$$T^1_{\max} = \max_{i:\delta_i=1}(T_i) \to \tau_0 < \tau_H(x) \text{ in probability as } n \to \infty.$$

Therefore, for n large enough, $T^1_{\max} \leq \tau_0 < \tau_H(x)$ for all $x \in I_\delta$ and taking $b = \tau_0$ we can apply the results considering $t = T^1_{\max}$.

Lemma 7 *Under assumptions (A1), (A2), (A3), (A4), (A8) and (A10) and if* $nh^5/\ln n = O(1)$ *and* $\ln n/(nh) \to 0$*, then the incidence estimator satisfies:*

$$1 - \hat{p}_h(x) = \exp\left(-\widehat{\Lambda}_h(T^1_{\max}|x)\right) + R_n(x), \text{ for all } x \in I$$

with

$$\sup_{x\in I}|R_n(x)| = O\left((nh)^{-1}\right) \; a.s.$$

Proof: The incidence estimator is equal to:

$$1-\hat{p}_h(x) = 1-\widehat{F}_h(T^1_{\max}|x),$$

where $\hat{S}_h(\cdot|x) = 1-\hat{F}_h(\cdot|x)$ is the Beran estimator in (2). The result can be found for $(t,x)\in[a,b]\times I_\delta$ in the proof of Theorem 2 in Iglesias-Pérez and González-Manteiga (1999) when the data are subject to random left truncation and right censorship. González-Manteiga and Cadarso-Suárez (1994) proved a similar result under right random censoring with fixed design on the covariate. □

Lemma 8 *Under assumptions (A1)-(A10) for $x\in I$, if $\tau_0 < \tau_G(x)$ for all $x\in I$ and if $nh^5/\ln n = O(1)$, $\ln n/(nh)\to 0$, then*

$$\widehat{\Lambda}_h(T^1_{\max}|x) - \Lambda(T^1_{\max}|x) = \sum_{i=1}^n \tilde{B}_{hi}(x)\xi\left(T_i,\delta_i,x\right) + \tilde{R}_n(x),$$

with $\tilde{B}_{hi}$ in (7), ξ in (8) and

$$\sup_{x\in I}\left|\tilde{R}_n(x)\right| = O\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \; a.s.$$

Proof: From Theorem 2(b) of Iglesias-Pérez and González-Manteiga (1999) when the data are subject to random left truncation and right censorship, and from Theorem 2.2 of González-Manteiga and Cadarso-Suárez (1994) with a non-random covariate using Gasser-Müller weights, it follows that

$$\begin{aligned}\widehat{\Lambda}_h(T^1_{\max}|x) - \Lambda(T^1_{\max}|x) = &\sum_{i=1}^n \tilde{B}_{hi}(x)\xi\left(T_i,\delta_i,x\right) + \sum_{i=1}^n (B_{hi}(x)-\tilde{B}_{hi}(x))\xi\left(T_i,\delta_i,x\right)\\ &+\sum_{i=1}^n B_{hi}(x)\left(\tilde{\xi}\left(T_i,\delta_i,x\right)-\xi\left(T_i,\delta_i,x\right)\right) + \tilde{\tilde{R}}_n(x), \qquad \text{(A.11b)}\end{aligned}$$

with ξ in (8),

$$\tilde{\xi}\left(T_i,\delta_i,x\right) = \frac{I(\delta_i=1)}{1-H(T_i|x)} - \int_0^{T^1_{max}} \frac{I(t<T_i)}{(1-H(t|x))^2} dH^1(t|x)$$

and

$$\sup_{x\in I}\left|\tilde{\tilde{R}}_n(x)\right| = O\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \; a.s.$$

Note that

$$\left|\tilde{\xi}\left(T_i,\delta_i,x\right)-\xi\left(T_i,\delta_i,x\right)\right| \le \int_{T^1_{\max}}^{\tau_0} \frac{dH^1(t|x)}{(1-H(t^-|x))^2} \quad \text{for all } i=1,\ldots,n.$$

Then, under assumption (A11) and using Lemma 6, it is easy to prove that for a sequence of bandwidths satisfying $nh^5(\ln n)^{-1} = O(1)$, the third term in (A.11b) is,

$$\sup_{x\in I}\left|\sum_{i=1}^{n} B_{hi}(x)\left(\tilde{\xi}\left(T_i,\delta_i,x\right)-\xi\left(T_i,\delta_i,x\right)\right)\right| = o\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) a.s.$$

For the second term in (A.11b), it is important to note that:

$$\sum_{i=1}^{n}(B_{hi}(x)-\tilde{B}_{hi}(x))\xi(T_i,\delta_i,x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\xi(T_i,\delta_i,x)\frac{m(x)-\hat{m}_h(x)}{\hat{m}_h(x)m(x)}$$

with $\hat{m}_h(x)$ the Parzen-Rosenblatt estimator of $m(x)$. Using Theorem 3.3 of Arcones (1997), standard bias and variance calculations and Taylor expansions lead to

$$\sup_{x\in I}\left|\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\xi(T_i,\delta_i,x)\right| = O\left(h^2+\sqrt{\frac{\ln\ln n}{nh}}\right) \text{a.s.}$$

Using again Theorem 3.3 of Arcones (1997), it is easy to prove that:

$$\sup_{x\in I}\left|\frac{m(x)-\hat{m}_h(x)}{\hat{m}_h(x)m(x)}\right| = O\left(h^2+\sqrt{\frac{\ln\ln n}{nh}}\right) \text{ a.s.}$$

Therefore,

$$\sup_{x\in I}\left|\sum_{i=1}^{n}(B_{hi}(x)-\tilde{B}_{hi}(x))\xi(T_i,\delta_i,x)\right| = O\left(\left(h^2+\sqrt{\frac{\ln\ln n}{nh}}\right)^2\right) \text{ a.s.}$$

For a sequence of bandwidths satisfying $nh^5(\ln n)^{-1} = O(1)$, it is immediate to prove that

$$\sup_{x\in I}\left|\sum_{i=1}^{n}(B_{hi}(x)-\tilde{B}_{hi}(x))\xi(T_i,\delta_i,x)\right| = O\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \text{ a.s.}$$

This completes the proof. □

Lemma 9 *Under assumptions (A1)-(A10) if* $nh^5/\ln n = O(1)$, $\ln n/(nh) \to 0$ *and if* $\tau_0 < \tau_G(x) = \tau_H(x)$, *then*

$$\sup_{x\in I}\left|\widehat{\Lambda}_h\left(T^1_{\max}|x\right)-\Lambda\left(T^1_{\max}|x\right)\right| = O\left(\left(\frac{\ln n}{nh}\right)^{1/2}\right) \ a.s.$$

Proof: The proof of the equivalent result for a fixed $t \in [a,b]$ is within that of Theorem 2(c) in Iglesias-Pérez and González-Manteiga (1999). For the uniform strong consistency of the Beran estimator $\hat{F}_h(t|x)$, see also Dabrowska (1989). □

Proof of Theorem 3. The incidence estimator can be split into the following terms:

$$\begin{aligned}
&(1-\hat{p}_h(x)) - (1-p(x)) \\
&= \hat{S}_h(T^1_{\max}|x) - (1-p(x)) \\
&= \exp\left[-\widehat{\Lambda}_h(T^1_{\max}|x)\right] - \exp\left[-\Lambda(T^1_{\max}|x)\right] + R_2(x) + R_3(x), \qquad \text{(A.12)}
\end{aligned}$$

with

$$\begin{aligned}
R_2(x) &= \hat{S}_h(T^1_{\max}|x) - \exp\left[-\widehat{\Lambda}_h(T^1_{\max}|x)\right], \\
R_3(x) &= S(T^1_{\max}|x) - (1-p(x)).
\end{aligned}$$

To the first term of (A.12) we apply a Taylor expansion of the function $\exp(y)$ around the value $y = -\Lambda(T^1_{\max}|x)$:

$$\begin{aligned}
&\exp\left[-\widehat{\Lambda}_h(T^1_{\max}|x)\right] - \exp\left[-\Lambda(T^1_{\max}|x)\right] \\
&= -\exp\left[-\Lambda(T^1_{\max}|x)\right]\left(\widehat{\Lambda}_h(T^1_{\max}|x) - \Lambda(T^1_{\max}|x)\right) + R_1(x),
\end{aligned}$$

with

$$R_1(x) = \frac{1}{2}\exp\left[-\Lambda^*(T^1_{\max}|x)\right]\left(\widehat{\Lambda}_h(T^1_{\max}|x) - \Lambda(T^1_{\max}|x)\right)^2$$

and $\Lambda^*(T^1_{\max}|x) = \eta_n(x)$ a value between $\widehat{\Lambda}_h(T^1_{\max}|x)$ and $\Lambda(T^1_{\max}|x)$. Now, adding and substracting $1-p(x)$, and bearing in mind that $S(T^1_{\max}|x) = \exp[-\Lambda(T^1_{\max})|x]$,

$$\begin{aligned}
&\exp\left[-\widehat{\Lambda}_h(T^1_{\max}x)\right] - \exp\left[-\Lambda(T^1_{\max}|x)\right] \\
&= (1-p(x))\left(\widehat{\Lambda}_h(T^1_{\max}|x) - \Lambda(T^1_{\max}|x)\right) + R_1(x) + R_4(x), \qquad \text{(A.13)}
\end{aligned}$$

where

$$R_4(x) = \left[S\left(T^1_{\max}|x\right) - (1-p(x))\right]\left(\widehat{\Lambda}_h(T^1_{\max}|x) - \Lambda(T^1_{\max}|x)\right).$$

Now, inserting (A.13) in (A.12), we have:

$$\begin{aligned}
&(1-\hat{p}_h(x)) - (1-p(x)) \\
&= (1-p(x))\left(\widehat{\Lambda}_h(T^1_{\max}|x) - \Lambda(T^1_{\max}|x)\right) + R_1(x) + R_2(x) + R_3(x) + R_4(x). \qquad \text{(A.14)}
\end{aligned}$$

The iid representation of $1-\hat{p}_h(x)$ now follows from Lemma 8.

Let us study the remainder terms in (A.14) starting with $R_1(x)$. Taking into account that $\exp[-\Lambda^*(T^1_{\max}|x)]$ is bounded for all $x \in I$, and applying Lemma 9, we have

$$\sup_{x\in I}|R_1(x)| = O\left(\frac{\ln n}{nh}\right) \quad \text{a.s.}$$

Regarding $R_2(x)$, directly from Lemma 7 and using $\ln n/(nh) \to 0$ we obtain:

$$\sup_{x\in I}|R_2(x)| = O\left((nh)^{-1}\right) = o\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \quad \text{a.s.}$$

Focusing on $R_3(x)$, note that it can be bounded as follows:

$$\begin{aligned}
\sup_{x\in I}\left|R_3(x)\right| &= \sup_{x\in I}\left|S(T^1_{\max}|x) - (1-p(x))\right| \\
&= \sup_{x\in I}\left|\left[(1-p(x)) + p(x)S_0(T^1_{\max}|x)\right] - (1-p(x))\right| \\
&= \sup_{x\in I}\left|p(x)S_0(T^1_{\max}|x)\right| \leq \sup_{x\in I}\left|S_0(T^1_{\max}|x)\right| = \sup_{x\in I}\left|S_0(T^1_{\max}|x) - S_0(\tau_0|x)\right| \\
&\leq \sup_{x\in I}|(T^1_{\max}-\tau_0)S_0'(\tau_n|x)|, \qquad \text{(A.15)}
\end{aligned}$$

with $\tau_n \in [T^1_{\max}, \tau_0]$. From condition (A5), that implies that there exists some $\lambda > 0$ such that $\sup_{(t,x)\in[a,b]\times I}|S_0'(t|x)| \leq \lambda$, and using (A.4) and (A.15) for a sequence of bandwidths verifying $nh^5(\ln n)^{-1} = O(1)$ we have that:

$$\sup_{x\in I}|R_3(x)| = o\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \text{ a.s.}$$

Finally, from Lemma 9, the term R_4 is negligible with respect to R_3, and therefore:

$$\sup_{x\in I}|R_4(x)| = o\left(\left(\frac{\ln n}{nh}\right)^{3/4}\right) \text{ a.s.}$$

This completes the proof. □

References

[1] Akritas, M. (1986). Bootstrapping the Kaplan-Meier estimator. Journal of the American Statistical Association, 81, 1032–1039.

[2] Arcones, M. A. (1997). The law of the iterated logarithm for a triangular array of empirical processes. Electronic Journal of Probability, 2, 1–39.

[3] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

[4] Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society, Series B, 11, 15–53.

[5] Cai, C.; Zou Y.; Peng, Y. and Zhang, J. (2012). smcure: Fit Semiparametric Mixture Cure Models. R package version 2.0. http://CRAN.R-project.org/package=smcure

[6] Cantor, A. B. and Shuster J. J. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data. Statistics in Medicine, 11, 931–937.

[7] Cao, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. The Annals of Statistics, 19, 2226–2231.

[8] Cao, R. and González-Manteiga, W. (1993). Bootstrap methods in regression smoothing. Journal of Nonparametric Statistics, 2, 379–388.

[9] Cao, R.; Janssen, P. and Veraverbeke, N. (2001). Relative density estimation and local bandwidth selection for censored data. Computational Statistics and Data Analysis, 36, 497–510.

[10] Chen, K.; Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. Biometrika, 89, 659–668.

[11] Chen, M. H.; Ibrahim, J. G. and Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. Journal of the American Statistical Association, 94, 909–919.

[12] Dabrowska, D. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. The Annals of Statistics, 17, 1157–1167.

[13] Dabrowska, D. (1992). Variable bandwidth conditional Kaplan-Meier estimate. The Scandinavian Journal of Statistics, 19, 351–361.

[14] Denham, J. W.; Denham, E.; Dear, K. B. and Hudson, G. V. (1996). The follicular non-Hodgkin's Lymphomas - I. The possibility of cure. The European Journal of Cancer, 32, 470–479.

[15] Efron, B. (1979). Bootstrap methods: Another look at the jackniffe. The Annals of Statistics, 7, 1–26.

[16] Efron, B. (1981). Censored data and the bootstrap. Journal of the American Statistical Association, 76, 312–319.

[17] Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. Biometrics, 38, 1041–1046.

[18] Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk? Canadian Journal of Statistics, 14, 257–262.

[19] Ghitany, M. E.; Maller, R. A. and Zhou, S. (1994). Exponential mixture models with long-term survivors and covariates. Journal of Multivariate Analysis, 49, 218–241.

[20] González-Manteiga, W. and Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. Journal of Nonparametric Statistics, 4, 65–78.

[21] Haybittle, J. L. (1959). The estimation of the proportion of patients cured after treatment for cancer of the breast. British Journal of Radiology, 32, 725–733.

[22] Haybittle, J. L. (1965). A two-parameter model for the survival curve of treated cancer patients. Journal of the American Statistical Association, 60, 16–26.

[23] Horvath, L. and Yandell, B. S. (1987). Convergence Rates for the Bootstrapped Product-Limit Process. The Annals of Statistics, 15, 1155–1173.

[24] Iglesias-Pérez M. C. and González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. Journal of Nonparametric Statistics, 10, 213–244.

[25] Jones, D. R.; Powles, R. L.; Machin, D. and Sylvester, R. J. (1981). On estimating the proportion of cured patients in clinical studies. Biometrie-Praximetrie, 21, 1–11.

[26] Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. Biometrika, 79, 531–541.

[27] Lai, T. L. and Wang, J. Q. (1993). Edgeworth expansions for symmetric statistics with applications to bootstrap methods. Statistica Sinica, 3, 517–542.

[28] Laska, E. M. and Meisner, M. J. (1992). Nonparametric estimation and testing in a cure model. Biometrics, 48, 1223–1234.

[29] Li, C. and Taylor, J. M. G. (2002). A semi-parametric accelerated failure time cure model. Statistics in Medicine, 21, 3235–3247.

[30] Li, G. and Datta, S. (2001). A bootstrap approach to nonparametric regression for right censored data. Annals of the Institute of Statistical Mathematics, 53, 708–729.

[31] Lo, S. H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. Probability Theory and Related Fields, 71, 455–465.

[32] Maller, R. A. and Zhou, S. (1992). Estimating the proportion of immunes in a censored sample. Biometrika, 79, 731–739.

[33] Maller, R. A. and Zhou, S. (1994). Testing for sufficient follow-up and outliers in survival data. Journal of the American Statistical Association, 89, 1499–1506.

[34] Maller, R. A. and Zhou, S. (1996). Survival analysis with long-term survivors. Chichester, U.K.:Wiley.

[35] Miyamoto, Y.; Hayashi, N.; Sakamoto, Y.; Ohuchi, M.; Tokunagam, R.; Kurashige, J.; Hiyoshi, Y.; Baba, Y.; Iwagami, S.; Yoshida, N.; Yoshida, M. and Baba, H. (2015). Predictors of long-term survival in patients with stage IV colorectal cancer with multi-organ metastases: a single-center retrospective analysis. International Journal of Clinical Oncology, DOI 10.1007/s10147-015-0835-2.

[36] Peng, Y. and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. Biometrics, 56, 237–243.

[37] Peng Y.; Dear, K.B. and Denham, J. W. (1998). A generalized F mixture model for cure rate estimation. Statistics in Medicine, 17, 813–830.

[38] Reid, N. (1981). Estimating the median survival time. Biometrika, 68, 601–608.

[39] Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. Biometrics, 56, 227–236.

[40] Tsodikov, A. (1998). A cure model with time-changing risk factor: an application to the analysis of the secondary leukaemia. A report from the international database on Hodgkin's disease. Statistics in Medicine, 17, 27–40.

[41] Tsodikov, A. (2001). Estimation of survival based on proportional hazards when cure is a possibility. Mathematical and Computer Modelling, 33, 1227–1236.

[42] Tsodikov, A. (2003). Semiparametric models: A generalized self-consistency approach. Journal of the Royal Statistical Society: Series B, 65, 759–774.

[43] Van Keilegom, I. and Veraverbeke, N. (1997a). Estimation and bootstrap with censored data in fixed design nonparametric regression. Annals of the Institute of Statistical Mathematics, 49, 467–491.

[44] Van Keilegom, I. and Veraverbeke, N. (1997b). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. Communications in Statistics, Theory and Methods, 26, 853–869.

[45] Wang, L.; Du, P. and Lian, H. (2012). Two-component mixture cure rate model with spline estimated nonparametric components. Biometrics, 68, 726–735.

[46] Xu, J. and Peng, Y. (2014). Nonparametric cure rate estimation with covariates. The Canadian Journal of Statistics, 42, 1–17.

[47] Yakovlev, A. Y.; Cantor, A. B. and Shuster, J. J. (1994). Parametric versus nonparametric methods for estimating cure rates based on censored survival data. Statistics in Medicine, 13, 983–986.

[48] Yakovlev, A. and Tsodikov, A. (1996). Stochastic models of tumor latency and their biostatistical applications (Vol. 1). OECD Publishing.

[49] Yamaguchi, K. (1992). Accelerated failure-time regression model with a regression model of surviving fraction: an analysis of permanent employment in Japan. Journal of the American Statistical Association, 87, 284–292.

[50] Zeng, D.; Yin, G. and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. Journal of the American Statistical Association, 101, 670–684.

[51] Zhang, J. and Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. Statistics in Medicine, 26, 3157–3171.